

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2003, 23:54:27 ; Search time 3900 Seconds

(without alignments)
17058.722 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 2286

Sequence: 1 atgggatactactcgaatgcac.....taaaaagaacacgtttaa 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

```

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_cm:*
21: em_or:*
22: em_ov:*
23: em_par:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_iny:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pla:*
35: em_htg_rtd:*
36: em_htg_man:*
37: em_htg_vit:*
38: em_sy:*
39: em_htgc_hum:*
40: em_htgc_mus:*
41: em_htgc_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	6.3	86827	3	PFMAL3P5
2	137.2	6.0	86827	3	PFMAL3P5
3	136.6	6.0	4601	3	DMU11584
4	136.6	6.0	19517	3	DMU37541
5	136.2	6.0	14867	3	AE001398
6	136.2	6.0	104992	2	AC005504
7	136.2	6.0	169546	2	AC004157
8	134	5.9	67970	3	PFMAL1P3
9	132.4	5.8	177293	2	AC069525
10	130.2	5.7	133877	2	AC120883
11	130.2	5.7	67970	3	PFMAL1P3
12	129.6	5.7	172816	9	AC093899
13	127	5.6	133877	2	AC120883
14	125.8	5.5	181792	9	AC098822
15	122.6	5.4	176898	9	AC117569
16	120.2	5.3	14867	3	AE001398
17	120.2	5.3	96953	9	AC079621
18	120	5.2	2426	8	SDU49822
19	119.4	5.2	206059	2	AC127383
20	118.4	5.2	178670	9	AC104073
21	118.2	5.2	137889	9	AC073269
22	118	5.2	111861	9	AC069435
23	118	5.2	172724	2	AC098590
24	117.6	5.1	15421	3	PFMCOMP1RA
25	117.4	5.1	79018	9	AL627107
26	117.4	5.1	205130	2	AC105425
27	117.2	5.1	104992	2	AC005504
28	117.2	5.1	169546	2	AC004157
29	117.2	5.1	177293	2	AC069525
30	116.6	5.1	90487	9	AL592166
31	115.8	5.1	172816	9	AC093899
32	115.6	5.1	184838	9	AC090740
33	115.4	5.0	96953	9	AC079621
34	115.4	5.0	155106	9	AC104069
35	115.2	5.0	72243	9	AL731858
36	114.8	5.0	143331	9	AC091214
37	114.4	5.0	170143	2	AC048384
38	114	5.0	349980	6	AX344555
39	113.8	5.0	54618	9	AC104800
40	113.8	5.0	172724	2	AC098590
41	113.4	5.0	106650	9	AC007708
42	113.4	5.0	148075	2	AC110014
43	112.6	4.9	155106	9	AC104069
44	112.4	4.9	26565	2	AC119399
45	112.2	4.9	4601	3	DMU11584

ALIGNMENTS

```

RESULT 1
PFMAL3P5      86827 bp      DNA      linear      INV 04-MAY-2000
LOCUS        Plasmodium falciparum MAL3P5, complete sequence.
DEFINITION   AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
ACCESSION   AL010206 AL010210 AL139179
VERSION      AL034556.3 GI:7711064
KEYWORDS     HTG; centromere; CTRP protein; initiation factor E4;
SOURCE       Serine/threonine protein phosphatase.
ORGANISM     Plasmodium falciparum 3D7.
REFERENCE    1 (bases 1 to 86827) Basham,D., Brown,D., Chillingworth,T.,
AUTHORS      Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
              Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T.,

```

TITLE	Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holtroyd,S., Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J., Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A., Rajadream,M.A., Rutter,S., Skelton,J., Squares,R., Squares,S., Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and Barrell,B.G.
JOURNAL	The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum
MEDLINE	Nature 400 (6744), 532-538 (1999)
PUBMED	99376098
REFERENCE	10448835
AUTHORS	2 (bases 1 to 86827)
JOURNAL	Bowman,S., Skelton,J., Churcher,C., Lawson,D., Quail,M. and Barrell,B.
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 86827)
TITLE	Lawson,D., Bowman,S. and Barrell,B.
JOURNAL	Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT	On or before May 14, 2001 this sequence version replaced gi:2982535, gi:2982536, gi:2894454, gi:2982554, gi:2982562, gi:2894489, gi:2982572, gi:2982574, gi:4493931. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P-falciparum .
FEATURES	location/Qualifiers
source	1..86827
	/organism="Plasmodium falciparum 3D7"
	/strain="3D7"
	/db_xref="taxon:36329"
	/clone="MAL3P"
	/chromosome="3"
gene	join(322..603,826..1023,1197..1301,1458..2942)
	/gene="PFC0575w, MAL3P5.1"
	join(332..603,826..1023,1197..1301,1458..2942)
	/gene="PFC0575w, MAL3P5.1"
	/note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 669 aa, possible signal sequence, revised: added new exon 2"
	/codon_start=1
	/product="Hypothetical protein, PFC0575w"
	/protein_id="CA838969.2"
	/db_xref="GI:7711065"
	/db_xref="SPIREREBL:097258"
	/translation="MTKANYITITSCFILDPLDSCFHEILKKTKKNNNNKSTFFLRSPTLRKRRSRKIKVSKFRKQKRPDLFENLEKSGFLGFWENQYDQYIDVITSNNINLRKQVTKVHKKSNEFTKENIKQILHCVPKSIDKIIINNLSITIKFQMSNTVHSILDISEKVRKEDAEYVIALHFLFKIDENITLFSMIMHPFKSKQKVCISIDIKSKRRKKNSIYENLFICTLTLYFCMLIKYISHICJFEFFPFCFLYILERIEECVGDILRRKIKIERIYNLCECKRKIKEMKRAIKMEIKNKDDIVENYDIELLCFTMKLIERNNKNNITISNYDNNINDISIDKDMYNNPDIYNNINISIDEKIKDEFPNDDENLKLKDTYQFLPNDNITIKYIEEDQPLYNINDSNINDNNINNTKNSHKRTKDINDDDYERKEEDIVTOKNDIDYIKNTYIGNKSJLEEKNOFIDADEIFQNFISNVNLDQGVKSDNDEVTGSTEHLKKNKTINKYQDDELJONEMNNPTIKNNIDNISINSEINIIILKILNOSDEIDINTSLIERLRTVLMVYIORKIEYLKFKYQDINFOYPIIKNKRTYVLDLNLNGYKIVMSPDVNSLFEFKTKIDSIPENKOKNNQMEKQKSNKTYITMNNW"
misc_feature	604..609
	/gene="PFC0575w, MAL3P5.1"
	/note="predicted splice donor sequence for exon 1 of PFC0575w (revised)"
misc_feature	816..825
	/gene="PFC0575w, MAL3P5.1"
	/note="predicted splice acceptor sequence for exon 2 (revised of PFC0575w)"
misc_feature	1024..1029
	/gene="PFC0575w, MAL3P5.1"
	/note="predicted splice donor sequence for exon 2 of PFC0575w (revised)"
gene	complement(3354..6644)
	/gene="PFC0580C, MAL3P5.2"
	complement(3354..6644)
CDS	/gene="PFC0580C, MAL3P5.2"

	/note="PFC0580c (MAL3P5.2). Hypothetical protein, len: 1097 aa, possible signal sequence, predicted using hexaXon"
	/codon_start=1
	/product="Hypothetical protein, PFC0580c"
	/protein_id="CAB38366.1"
	/db_xref="GI:44939332"
	/db_xref="SPIRMBL:097257"
	/translation="MLGILRRKRVFVLLVSPSLFAYFLIKRKHQDNNNYETLLINNDIE KIKRIIRHNKCSYIPLFLNIAYDSYIYKNIILMWLFERKRRKRCDEYYITVMVK KRREKIKENFISDEQNLFNKFYIYEILFVSLKYGILSHLSLYLKLISPCVNIYF SLYYNKKLNDKNNHLEKRLKFKYQIINNEHQOAPRNHHNNNNNNKPLDINHSC KNNISYSSVSTNNNEKNNINIDKYNVHNHEFKRSTSYEDNCOHITLMMYILNORY DNICICLNTNNTNINYNININILIKYICNNMETILLNNHLEDKKTITNOKNTNYS LPKITYISFFPKKEENHITIDLEDDMMNNHLKRNDEKFTYNSNETHNNIYKISDY FEDHNSNSNNCSFNLKQOQTDNTRKHIIMGEKRYPMKRSOHEKNNNTGNINIEK DOCKOILKRIKIFLGNKRLIDTQILNELYVIRLIFESCLISIKRHHILKELKE PEKDKITIIYLSADYNNLRNRIILKRSKKEEENINSPASPLFSLKNIITFDEIG FSDRTIYANNQDETRKTTNNNDNNNDNCSNDHICSNNNHICSNNDNIGCN NNNNNNNNNNCSNNNNNCSNNNNCSNNKMLFEDPQDNKPDYKRRKKRRTI VBLAIYNNNIFDLKCKRKHOREDTINIKYIKKRPMTIYLNITIKNDOTFTEH NNNIYNGIKNNNITLKKKTNLFESSLISFYIKQIDYEYVAKRLIYALYDLIKRFR VFLKNNHINLKLKMKRAFHYIYNEDQFTLNYYHILHKKNIKHILHKOQKDE IDIVFCKLYCMATNININIFSYIHAKVNECVYRFFKAYNILLFENYIENEKENY YKNRKTIYANNINISNDLYPNISYKNIYQNKRRKNSFTFYLKNDHILFL LYTQILQCCDIFSYIKKYKNEKRNPLNLYIELHYVSEKKKKKKFPSPSSP SYSDMMVSEFFSYFFESLSYLFLIFLPHDMASITFEKITLTSGLPTYYISLYNNI MYSCPKR"
gene	join(7669. 7717, 7800. 7829, 7912. 7940, 8064. 8106, 8302. 8368, 8568. 8641, 8713. 8812, 8924. 9006, 9122. 9258 9369. 9505, 9613. 9838)
	/gene="PFC0581w"
CDS	join(7669. 7717, 7800. 7829, 7912. 7940, 8064. 8106, 8302. 8368, 8568. 8641, 8713. 8812, 8924. 9006, 9122. 9258 9369. 9505, 9613. 9838)
	/gene="PFC0581w"
	/note="PFC0581w, hypothetical protein, len: 324 aa, revised: new gene prediction, splicing prediction using tentative"
	/codon_start=1
	/product="hypothetical protein, PFC0581w"
	/protein_id="CAB90285.1"
	/db_xref="GI:7711066"
	/translation="MKKKKKHYPSIFKPVNDKKNRYVLCSDYIRIINMYGLHIFRL QKIHIFIAEQIALSSFFESHKKNRYVTKEDIDGLEKQMFNTDILLESIKQDNDS NDITVYVYAMKNNKQDLITTSYGNKRLIDVYNSFQMAQSPFYLINIKFSHRMSPPA LKDVADEIVSKNNKSPSALSNDNSVTKYKYIVLTDLNIESTKYNFASGVKKYV TLKREKKKNNRILSKREKYPNKCVMQDKKRRKHITTFVILNIFPLISLSHRHDSV ONFLKERNKNSDKLQDIDDEDEKRYFDEETILKEAKKKSEYDKDDEEL"
misc_feature	7790. 7799
	/gene="PFC0581w"
	/note="potential splice acceptor sequence for exon 2 of PFC0581w"
misc_feature	7830. 7835
	/gene="PFC0581w"
	/note="potential splice donor sequence for exon 2 of PFC0581w"
misc_feature	7901. 7911
	/gene="PFC0581w"
	/note="potential splice acceptor sequence for exon 3 of PFC0581w"
misc_feature	7941. 7947
	/gene="PFC0581w"
	/note="potential splice donor sequence for exon 3 of PFC0581w"
misc_feature	8055. 8063
	/gene="PFC0581w"
	/note="potential splice acceptor sequence for exon 4 of PFC0581w"
misc_feature	8107. 8112
	/gene="PFC0581w"
	/note="potential splice donor sequence for exon 4 of PFC0581w"

	misc_feature	8293.. 8301	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 5 of PFC0581w"
	misc_feature	8369.. 8374	/gene="PFC0581w"	/note="potential splice donor sequence for exon 5 of PFC0581w"
	misc_feature	8539.. 8567	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 6 of PFC0581w"
	misc_feature	8642.. 8647	/gene="PFC0581w"	/note="potential splice donor sequence for exon 6 of PFC0581w"
	misc_feature	8704.. 8712	/gene="PFC0581w"	/note="potential splice donor sequence for exon 6 of PFC0581w"
	misc_feature	8813.. 8818	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 7 of PFC0581w"
	misc_feature	8913.. 8923	/gene="PFC0581w"	/note="potential splice donor sequence for exon 7 of PFC0581w"
	misc_feature	9007.. 9012	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 8 of PFC0581w"
	misc_feature	9112.. 9121	/gene="PFC0581w"	/note="potential splice donor sequence for exon 8 of PFC0581w"
	misc_feature	9259.. 9264	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 9 of PFC0581w"
	misc_feature	9361.. 9368		
	Query Match	6.3%;	Score 143;	DB 3; Length 86627;
	Best Local Similarity	45.0%;	Pred. No. 7.9e-11;	
	Matches 814;	Conservative 0;	Mismatches 960;	Indels 35; Gaps 6;
OY	62 ATATGAAACATTAGCGTGCATTAAGTTAAGCGGACGAATAAACCAACAATATATAAA	121		
Db	39131 AAATAATTAATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39190		
OY	122 ACCTACTATTCGTAGCAATGGTGCTGTTAATGGTATGAAGAATGATGGAAGTAT	181		
Db	39191 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39247		
OY	182 TTCCGAGTTCCGAGCOTTTTTATGACGATTTTACCTATGCTAAGCAAGCGATACAA	241		
Db	39248 TATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39307		
OY	242 TCACAAAATCAAAACGATATATCATGATGCCACATAGCTGTAATAAATAGCATATC	301		
Db	39308 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39367		
OY	302 ATTTTTACTTAAGCACCATGCGTTATTTTGATTAATTAATTAATTAATTAATTAAT	361		
Db	39368 TATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39427		
OY	362 TAAATCTGCGAGAGAAATGAACACACATTAATTAATTAATTAATTAATTAATTAATTAAT	421		
Db	39428 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39487		
OY	422 AAAATCAAATGTAATTTTAGAAAAACGTTAAATCTTCAATCAATTTAGATTAAACA	481		
Db	39488 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39547		

[illegible]

[illegible]

misc_feature	8302.. 8368,8568.. 8641,8713.. 8812,8924.. 9006,9122.. 9255,9369.. 9505,9613.. 9838) /gene="PFC0581w" /note="PFC0581w, hypothetical protein, len: 324 aa, revised: new gene prediction, splicing prediction very tentative"
misc_feature	9790.. 9799 /codon_start=1 /product="hypothetical protein, PFC0581w" /protein_id="CA890285.1" /db_xref="GI:7711066"
misc_feature	7830.. 7835 /translation="MKKKKKHYFISIKFVNDKKNKYLCSDKDIIRINMIGLHIFRLLQKIHIFAEQJASPEFSKHYATKEDIJDGEEKMPTPLILSESIRKODPFSNDITYTAKKNKMDOLLTYSNKKIDTVNKSFCOMASPEYIFLNKFSRMSPPVALKVDVEIDFARKNKNSFSAISNDNSVTRKYIVDLTDNIISETYNFRASVKVYTLKEKKKINRLLSKERYPNMOMDKERHIIITFYINFLILSLSHRHDSVQNFKEKKNSDKLQDDIDDEDEKIFDEILTEAKKSEYDKDEEL"
misc_feature	7790.. 7799 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 2 of PFC0581w"
misc_feature	7830.. 7835 /gene="PFC0581w" /note="potential splice donor sequence for exon 2 of PFC0581w"
misc_feature	7901.. 7911 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 3 of PFC0581w"
misc_feature	7941.. 7947 /gene="PFC0581w" /note="potential splice donor sequence for exon 3 of PFC0581w"
misc_feature	8055.. 8063 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 4 of PFC0581w"
misc_feature	8107.. 8112 /gene="PFC0581w" /note="potential splice donor sequence for exon 4 of PFC0581w"
misc_feature	8293.. 8301 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 5 of PFC0581w"
misc_feature	8369.. 8374 /gene="PFC0581w" /note="potential splice donor sequence for exon 5 of PFC0581w"
misc_feature	8559.. 8567 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 6 of PFC0581w"
misc_feature	8642.. 8647 /gene="PFC0581w" /note="potential splice donor sequence for exon 6 of PFC0581w"
misc_feature	8704.. 8712 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 7 of PFC0581w"
misc_feature	8813.. 8818 /gene="PFC0581w" /note="potential splice donor sequence for exon 7 of PFC0581w"
misc_feature	8913.. 8923 /gene="PFC0581w" /note="potential splice acceptor sequence for exon 8 of PFC0581w"
misc_feature	9007.. 9012 /gene="PFC0581w" /note="potential splice donor sequence for exon 8 of PFC0581w"
misc_feature	9112.. 9121

Query Match	6.0%;	Score 137.2;	DB 3;	Length 86627;
Best Local Similarity	45.3%;	Pred. No. 4,7e-10;		
Matches 965;	Conservative 0;	Mismatches 1123;	Indels 44;	Gaps 11
misc_feature	/gene="PFC0581w" /note="potential splice acceptor sequence for exon 9 of PFC0581w" 9259..9264 /gene="PFC0581w" /note="potential splice donor sequence for exon 9 of PFC0581w" PFC0581w" 9361..9368			
misc_feature	9361..9368			
Query 79	TACATTAAGCTATACGACGCAAAAAAACACCAATATATAAAGCTACTATCTGTA	138		
Db 40793	TAAATTAAT	40734		
QY 139	GCAATGGTGGTTAAATGGTTATGAANTGATGTTGAAGATTTCCGAGTTTCGATCT	198		
Db 40733	AAAAATATATAAAAT	40674		
QY 199	TTTATGACGATTTTATACGTATGTCAGACGATACCAATCAACAATCAAAAAACA	258		
Db 40673	ATGTAATTAAT	40614		
QY 259	GATATATCATGATTCGACATACCTGTAATTAATACGATATCATTTTCTACTAAGAC	318		
Db 40613	AAT	40570		
QY 319	ACCATGGCTTATTTTGATATATATATACGCGAAATATATATATATATATATATAT	378		
Db 40569	AATTAAT	40510		
QY 379	AATGACACACATTAATAAATGAAGAGCGCTACTATTTTACCCAAAAATCAAAATGAT	438		
Db 40509	TATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40451		
QY 439	TTTAAAAACGCTTAAATCTTCATCATATTTAGATTTAACATGTTTAAATGGTTT	498		
Db 40450	AAAAATTAATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40391		
QY 499	AAATTAATATATATGATTAACCTTATGAACCAATACATCAATGCAACATTAAGTAG	558		
Db 40390	AAAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40331		
QY 559	AAATTAATGATGGTGTATTTTACAGATTCACACTTAACACATTTTAATTAACG	618		
Db 40330	ATAATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40271		
QY 619	ATTATTGAAGAATGATGATGAATGATGAGAGACCCATACATGCTGCGAATGCT	678		
Db 40270	ATTATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40211		
QY 679	TTTGCAAAACTCACACCTGCAACA--ACTTACATACATTCATTAATGACGATTAAT	736		
Db 40210	TACGGTGTGTAATAATATATATATATATATATATATATATATATATATATAT	40151		
QY 737	GTATGGCATATTCATTAATAGTATATATTTCCAAATTTTGACTATACAAATTAACAT	796		
Db 40150	AAATTAATATATGTTTCTAATAAATAAATAATTAATAAATACTATATATATTAAT	40091		
QY 797	TTTCTTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	854		
Db 40090	AATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	40031		
QY 855	CAACCAATATCAACATATTAATAATCTTATACCATTAATCAATTCACATATGATTT	914		
Db 40030	TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39971		
QY 915	TTATGACTATATTAATCATTTCTTCGTGGTGTTTAAATATGTATTAACCAATATCAT	974		
Db 39970	ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39911		

CDS
3083..3767
/note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except=(pos:3767,aa:TERM)
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="WSTWANGLDQSPALMEQOLIEFHDAHLILVMTIVGVLTLEM
LEFNNVNRFLHGLIMIMVILPAIILFLAPSLRLYLLEINPEVSILSGH
OWMSYEYEDENNIEFDSYMIPTNEMLDGFLDNDRVVLPNNSQRIITVADVI
HSMVPAVGKVDGTPGRNCTNFPIINRPLGYCGSCSEICGANHSFMDIVLESVNV
FIKWTSSNNS"
3768..3838
/product="tRNA-Lys"
3840..3906
/product="tRNA-Asp"
3907..4068
/codon_start=1
/transl_table=5
/product="AtPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPQWAFISWLLFLIFISITFLPFSINYSYMPNPSKSNELKNI
NLNSNMKW"
4062..4736
/codon_start=1
/transl_table=5
/product="AtPase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/translation="WMTNLFSECDPLAIPNFSLNLTFLGLIMPSIYWLMPSSYNI
MNSLITLHKPEKFLIPSGHNSGTFLIFSLILENMGSLPEYITSTSHLIT
LSLAPLHCPEFLYGMINTQMPAHVNOGPAIIMWCIETISNITRECTAVR
LRNMAGHLITLIGNTSSSYMLMFLMAQIALVLESAAVADIOSYRAVAVLSTL
YSSEVN"
4736..5524
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAC47816.1"
/db_xref="GI:1166535"
/translation="WSTHSNHPFLVDSPRLTGAGAMTVSGNKKRFHODISLF
VGNITLITVQWEDVSREGYQGLHTYAVTIGLRWGNILFLISEVLEFSFWAF
FHSLSPALEIGASWPMGIISFNPOPLPLNTAILLAGSVITWANHSLMNSQOT
TQGLFTVLGIVETILQAYEYIEAFITADSIYGSTFPMATGFQIHVLIGTFLVY
GLRHLNHSKHHNGFEAAMVHFDVWVLELYITLYWGG"
5543..5607
/product="tRNA-Gly"
5608..5961
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/protein_id="AAC47817.1"
/db_xref="GI:1166536"
/translation="MFSITFIALLILITTYMFLASIIISKALIDREKSPPECFD
PMSSSLPFSRLRFITLIFLDVEIALILPILIMKISIMIMTITSLITLILILI
GLYHEWNGMLWNSN"
5981..6045
/product="tRNA-Ala"
6055..6118
/product="tRNA-Arg"
6119..6183
/product="tRNA-Asn"
6184..6251
/product="tRNA-Ser"

Query Match 6.0%; Score 136.6; DB 3; Length 19517;
Best Local Similarity 44.9%; Pred. No. 9.3e-10;
Matches 955; Conservative 0; Mismatches 1139; Indels 35; Gaps 10;
QY 147 TTGGTTAATGGTTAGAAATGATGTTGAAGTATTTCCGAGTTTCGAATCTTTTATGA 206

Db 19401 TTTTTTTAAATTTTATGAAAAATTTTAAATTAATTTTTCATATATATATAT 19342
QY 207 CGCATTTTATACGTATGTGAAAAAGCGTGAATCAATCAATCAAAATCAAGATATTAT 266
Db 19341 TATATATAAATATTCATAATATATATATATATATATATATATATATATATAT 19282
QY 267 CATGATCGACATTAATCTGTAATTAATGATATATATATATATATATATATATATAT 326
Db 19281 ATATTAT 19222
QY 327 TTAATTTGTAATATATTAACAGCGAAATATATATATATATATATATATATATATAT 386
Db 19221 TTAATTCAT 19162
QY 387 CACAT-----TAAATGAAGAGCGTACTATTTTACCGAAATCAAAATGTATAT 438
Db 19161 TATTAATATTTAT 19102
QY 439 TTAGAAAAACGTTTAAATCTTCAATCAATTTAGATTTTAAATGATTTT 498
Db 19101 TTAATAAAAAATTTTTTTTTTAAATAATTTTTTAAATAATGAAATATATATAT 19042
QY 499 AATTTAT 558
Db 19041 ATATTTCAT 18984
QY 559 AATTTACTGATGCGTATTTTAAACAGACAGACACTTAAACAGATTTTAAATATACG 618
Db 18983 ATTTTAT 18924
QY 619 ATTTTGTAAAGAT 678
Db 18923 ATATTAT 18864
QY 679 TTTCGAAACCTACACCTGACACACTTACATCATTAAGACGTGATTTATAGT 738
Db 18863 CAT 18809
QY 739 ATGTGCCATATTCAT 798
Db 18808 AT 18750
QY 799 TCATTAAT 858
Db 18749 ACATTTTAAATTTTAAATTTTATATATATATATATATATATATATATATATAT 18690
QY 859 CATATCAGAT 913
Db 18689 ATTTAT 18630
QY 914 TTTATGACTAT 973
Db 18629 AATAAT 18570
QY 974 TAAACAACAT 1032
Db 18569 TTTTAAAT 18510
QY 1033 GTGATGTATATGAAAAATTCACACATGTTATATATATATATATATATATATATATAT 1092
Db 18509 ATAAAT 18450
QY 1093 ACCTTAATCCCTCTTTTATAGTATGACATATATATATATATATATATATATATAT 1152
Db 18449 ATTTAAT 18390
QY 1153 GATGATTTAAGCAT 1212
Db 18389 TAAATTTAT 18330
QY 1213 GTAAATTACTAT 1272

RESULT 5
AE001398/c

/organismPlasmodium falciparum 307"
 /strain="3D7"
 /db_xref="taxon:36329"
 chromosome="2"
 complement(1570..2424)
 /gene="PF80490c"
 complement(1570..2424)
 /gene="PF80490c"
 /note="Predicted by Glimmer"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAC11887.1"
 /db_xref="GI:3845198"
 /translation="MKKRIENIDYLSCELDVDVREKSGKNSLLKSSSTRKSDYKRS
 SIEFKKSDHKRKSQSRGSGRSGFIRKSGSRKPYNNRLINKNYNYKGRNFHNSR
 DNEKRTSEFSRPFDRKSGSEFKKPIISRNKSSVYSKSGNSNMKGSPNKAPTSS
 TVYRKLNNYTVGAPKPKNNNLISLYKRNKPLPLNKRSPVATIGTSSVPRRIKK
 GLVKSGLSKTRKSTSSSRKPLNKFLEKIKITVYSLKKIPSELEKNTFVNLPESL
 NNAETPKK"
 11241..14606
 /gene="PF80495w"
 11241..14606
 /gene="PF80495w"
 /note="Predicted by Glimmer"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAC11888.1"
 /db_xref="GI:3845199"
 /translation="MFALKNNIRREGFVNICEFSYLKLYLKSNSFVYVNLNRYETNEKR
 TNKTIYKRSKQSLEFGLKLNHKLPLFNKDKRYCARYECISAKYKYLDEYKRC
 ENYLCIDIDISYTFPELCKPTPTDNYFLVKNNDKVLKNEIYFKNDIFHAR
 EILKIDKINISYTHENGSELHPLIYPRNNKDKNLILEYVYFVNYDHAHLEP
 HEIKYLEIFKRYLNNSNLTPEKRLIOEMENLYPREIKENKYNLIKKMKREIYK
 PFAEFHNVQHDINENLITRIYVYSLDIDINKKMLNNNLINENIYESTLLK
 PFYOTLLKCKYQNDMTYKLYKATGHILCDTKNLEPCSDIDYSTILNSLNK
 FLNKLIDKNTLPEKRIKTEENAVLIHNDQNYSNKENIKDIIQKRIKEY
 LFNIDMKFSLNLEPGRIKTEENAVLIHNDQNYSNKENIKDIIQKRIKEY
 IYFVMEVQOFRKLDKSDLSIKLSTNFVKNINVSXDFELFNISCIILNPLV
 NRSVYAKQDTYLIINDLSFYKIKNDPRKKNFELSSSKMLICNLISNLY
 INKILHEEDHPQDQVSLTEFLNLEPDKIHFHYIYLMCHVATRYEKNCLN
 RYEDHISTLLTCKSDQYFLENSNDRCRKLHLKYNIIIDLIKNVNLISID
 NISKIPLISNSKYATEVNDMLBLSQSEPKYKTSKSGKIHMDNMLDNNSSE
 KYERRITPEKKNLTNKLNIIECLIKLPLILKKKTYLTKQSCSPINKENTL
 KILYIYLNLYLMEVGYCEMLERYSKQENLSTVYKAKVNEKRPDLICHSIE
 QDLEKSNYVLEYEDKLNINSESRNLINENLNDPDIIDEIKKRYKLNNTLTKH
 INVALYENKSNNSNGNSNLTLDKDKKNNHNNYNDINKNENKIDCKQNGNCIN

CKDLVNDIINIEGFLKMKRRKFLFOLYMLCNITKFKRRVSSSLFHMDEKLIK
DMNKYLCEWYKINECEAFYTDIYLFKER"

BASE COUNT 6284 a 1019 c 1106 g 6456 t
ORIGIN

Query Match 6.0%; Score 136.2; DB 3; Length 14867;
Best Local Similarity 44.2%; Pred. No. 1.2e-09;
Matches 967; Conservative 0; Mismatches 1198; Indels 22; Gaps 9;

QY 24 ATATCGTAACATGAGCGCATGATTTTATACGGGTATAGAAACCTTAGCGGACAA 83
DB 9362 ATATATTAATTAATTTTAAAAATATATATTAATTAATTAATTAATTAATTA 9303
QY 84 TAAAGTTAAGCGAGAAAAAACCACCAATATAAAAAGCTTACTTATCTGTAGCAAT 143
DB 9302 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9243
QY 144 TGGTGTGTTAATGCTTAAGAAATTCATGTCGAGATTTCCGAGTTTGGAACTTTTA 203
DB 9242 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9183
QY 204 TGACGATTTTATACGTATGTGMAAGAGCTGATACATCAACAAATCAAAACAGATAT 263
DB 9182 TAACAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9123
QY 264 TATCATGATTCGACATPACTGTATATTAATGATTAATTCATTTTCTTAAGACACCAT 323
DB 9122 TTTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9064
QY 324 GCGTATTTTGAATATATTAACGCGAAATATATTTTAAATTCGACAGAAATATGA 383
DB 9063 TAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9004
QY 384 ACACACATTAATAAATGAAGAGGCTACTATTTAGCCAAAATCAAAATGTAATTTTGA 443
DB 9003 TTTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8944
QY 444 AAAAGCGTTAAATCTTCATCAATTTAGATTAACAAAGTTTAAAGSTTTAAAT 503
DB 8943 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8886
QY 504 TAATATATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 563
DB 8885 ATAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8826
QY 564 ACTTGATGGTGTATTTAAGCAATCAGACTTAAACGATTTTAATATATCGATTTT 623
DB 8825 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8765
QY 624 TGAATAAGATATGATTA--TGAATGATAGTGAAGCCTATGACTATCTGGAATGTTT 681
DB 8765 AAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8706
QY 682 GCAAACTCAGCCTGACACACTTACATCATTCATTAAG--ACGTGATTTATATAGCTAT 740
DB 8705 TATATGTTTAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8646
QY 741 GTGCGATATTCATTAATGATATATTTCCAAATTTTGACTATTAACAAATTAACATTTTC 800
DB 8645 TTATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8586
QY 801 ATGGAATATTAATGGAATCTTACTGGAATAATGAATGACAGCTTTTCAGTTTACTCAACCA 860
DB 8585 TAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8526
QY 861 ATATCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 920
DB 8525 TAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8466
QY 921 CTATATTAATCAATTCATGCTGTGTTAAATATGATATAGCAACCAATTAATTAACAA 980
DB 8465 TTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8406

QY 981 ACTAATGATGAGCCTTGTTTCTTATGACATCA--ATTCGAGTTATCCTTATGATG 1038
DB 8405 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8346
QY 1039 TATCATGAAAAATTCACACATGCTTATACCTTTTACGACACATATTCGACCAACCGTTA 1098
DB 8345 CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8286
QY 1099 ATCCCTACTTTTATGATGATGACATTAATTTTTCATTAATTAATTAATTAATTAATTAATTAATTA 1158
DB 8285 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8226
QY 1159 TTTAAGCATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1218
DB 8225 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8166
QY 1219 TACTTAATTAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1278
DB 8165 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8106
QY 1279 GACATTAACGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1338
DB 8105 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8046
QY 1339 GAATACCTTTCATGACGAGATTTATTTTTCAAACCTTTTATTAATAACCAAGGTAAAG 1398
DB 8045 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7987
QY 1399 TTTAAATAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1458
DB 7986 TTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7927
QY 1459 GAAACCCCTACTCAAAAGAGAGGTATGTTATCTTAAGTCGTTTAAATGATTAAT 1518
DB 7926 AAATA--AAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7870
QY 1519 GGCATTAACGCTTAATGCTGACATTTTAACTTAATTCGTTTGAATGATTAACATGACTA 1578
DB 7869 TAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7810
QY 1579 TACATATCATTAACGGTTACAAAAACACTGAACGTAAATATATTCCT--ACATTT 1635
DB 7809 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7750
QY 1636 GTACACATCAGCTTCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1695
DB 7749 ATATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7690
QY 1696 ATGACGACAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1755
DB 7689 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7630
QY 1756 CCGTTTGAACCCAGTTTATTCGACCGCATAGCCTTGGAATGGAATGGAATGGAATGGAATGGAATGGA 1815
DB 7629 TAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7570
QY 1816 GAACAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1875
DB 7569 TTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7510
QY 1876 AAGATTAATAATTCCTCTGCTGATACCGAAAAAGCGCTTGATTAACAGCGCTCATTTT 1935
DB 7509 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7450
QY 1936 GAAACCTTGTGACGTAACAAATTCCTGACGGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1995
DB 7449 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7396
QY 1996 TATTAATGACAAAGGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2055
DB 7395 -TTATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7337
QY 2056 GTATATGATGAATATTTTACTGATGAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2115

	Db	74098	ATTAATTATAAATAATTAATTATTTAAATTAAATTAATAATTAATTAATTAATTAATTTT	74157
OY	1327	ATATATGAATGTAATACCTTCATCGACGGCATATTAATTTTTCAAACTATTTATTA	A	1386
Db	74158	ATTGTG-----TTAATATTATTAATCTATTTAAATATTAATTAATTAATTAATTTAT	T	74212
OY	1387	ACACAAGTAGTAAAAAACAATAATCATATGACATCACCCTAGACTATCACACTCT	CT	1446
Db	74213	TAATTAATTTTTTTAAAATAAAATATATTTATTAATAATGTAATTAATAATTAAGA		74272
OY	1447	GATGATATCAAAGAACCCACTACTCAANTGAGAGGTTATGTAATCTAAAGTCGT--	--	1503
Dg	74273	AATTAATTATTAATAATATATGTATTAATTAATTAATTAATAATTAATTAATTAATA		74332
OY	1504	---TTAAATGAGATTAATGSCATACCTGCATCTAGCTTCACATTTAACTTAAT--	-CCGTTT	1559
Db	74333	TTAATTTTAAATTAATTAATGTTTATTTATTAATTAATAATTAATTAATTAATTAAT		74392
OY	1560	AGATGATTAACAATGACATCATATCATATCATACGGTTACAAAACACTGACATTAAT		1619
Db	74393	TAATATATTATTTAAATTAATTAATTAATAATTAATTAATGAATATTAATTTTAA	TA	74452
OY	1620	ATTATTTCTCATTTTGTCACATCACCGTTCAATGTATTAATTTGGTTCCTTCCA	ATA	1679
Db	74453	ATTATTAATTAATTAATTAATTTATTTATTTATTTAATTAATTAATTAATTAATTA		74512
OY	1680	CTTAACGGAAGTGAATGACGACATTTTATTAATTTGCGATGCTGATGTTGTAAT		1739
Db	74513	TTTATTAATAATATATATATATATATATATATATATATTTATTTATTTAAATTA	AT	74572
OY	1740	GAATTCGCTGTAAACCCCTATTGAACCCGATTTATTCGACCGGATAGCCTTAGTAA		1799
Db	74573	TAATTAATAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTT		74632
OY	1800	ATGGATATTTGAAACGACAGATGATGATGATGTTGTACGATCATAGAATATGCG		1859
Db	74633	TTAATTTATTTAAATTAATTAATAATAATAATTTATTTATTTATTTATTTAAATTA	TA	74692
OY	1860	ATA ATA 1862		
Dd	74693	ATA 74695		
RESULT 7	AC004157	169546 bp DNA linear HTG 12-AUG-2000		
LOCUS	Plasmidium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN			
DEFINITION	AC004157	AC004157		
ACCESSION	AC004157	GI:9797712		
VERSION	HTGS_PHASE1.			
KEYWORDS	Plasmidium falciparum.			
SOURCE	Plasmidium falciparum.			
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
REFERENCE	1 (bases 1 to 169546)			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Rowley,D., Mao,J., Tamaki,T.,			
TITLE	Kundi,O.B., Conway,A.B. and Davis,R.W.			
JOURNAL	Plasmodium falciparum 3D7 chromosome 12			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 169546)			
TITLE	Hyman,R.W., Qiu,F., Fung,E.L., Conway,A.B. and Davis,R.W.			
JOURNAL	Direct Submission			
COMMENT	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology			
	Center, Stanford University, 855 California Avenue, Palo Alto, CA			
	94304, USA			
	On Aug 12, 2000 this sequence version replaced gi:9810447.			
	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 2 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Gaps between the contigs are represented as			
	* runs of N, but the exact sizes of the gaps are unknown.			
	* This record will be updated with the finished sequence			
	* as soon as it is available and the accession number will			

[illegible]

QY 847 CAGTACTGACCAATATCAAGATTTAAATATCTTATACACATATTCATTCATGAT 906
 DB 85265 TTATATTAATAAATATTAATAAATTAATTAATTAAGATATAATATAATATATATA 85324
 QY 907 ATGAATTTTATGACTATAATTAATCATTCATCGTGGTGAATATATGATACACC 966
 DB 85325 TTATATTAATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85381
 QY 967 AATTAATTAACCAACTATATGATGACCGCTGTTTTCATATGACATCATTCGCTAT 1026
 DB 85382 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85441
 QY 1027 CCTATGATGATCATGAAAAATTCACATGCTATACCTTTTACGACACATATCA 1086
 DB 85442 ATATATTAATTAATTAATTAATAAATATATATATTAATTAATTAATTAATTAAT 85501
 QY 1087 GAACCAACGTTATCCCTACTTTTATAGATGATGACATTAATTTTCATTAATAGAT 1146
 DB 85502 ATTTTATTAATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 85561
 QY 1147 GATTAAGATGATTTTACGATGATTTTATTAATTAATTAATTAATTAATTAATTAAT 1206
 DB 85562 AATTAATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85621
 QY 1207 ATGATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1266
 DB 85622 ATTAAGATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85681
 QY 1267 AGATATGATCAACATATCGGGTATGATGATGATGATGATGATGATGATGATGATGAT 1326
 DB 85682 ATATATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85741
 QY 1327 ATATATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1386
 DB 85742 ATTTGT-----TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85796
 QY 1387 ACCAAGATGATTAATAAACAACATATGACATGACATGACATGACATGACATGACATGAT 1446
 DB 85797 TAATTTATTTTATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 85856
 QY 1447 GATGATATGACGACACACCATATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1503
 DB 85857 AATTAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85916
 QY 1504 ---TTAATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1559
 DB 85917 TATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85976
 QY 1560 AGATGATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1619
 DB 85977 TAATTTATTTTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 86036
 QY 1620 ATATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679
 DB 86037 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86096
 QY 1680 CTTAAGGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1739
 DB 86097 TTTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86156
 QY 1740 GAATCCGTTGTTAAACCTTATGAACCCGTTTATGACCCGATAGCCCTAGGTA 1799
 DB 86157 TAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86216
 QY 1800 ATGGGATTTGAAGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1859
 DB 86217 TTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86276
 QY 1860 ATA 1862
 DB 86277 ATA 86279

RESULT 8
 PFMALIP3
 LOCUS PFMALIP3 67970 bp DNA linear INV 15-DEC-1999
 DEFINITION Plasmodium falciparum MALIP3, complete sequence.
 ACCESSION AL031746
 VERSION AL031746.9 GI:6594243
 KEYWORDS HMG.
 SOURCE
 ORGANISM Plasmodium falciparum 3D7.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 67970)
 AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.
 AND Berrill, B.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
 The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA, UK
 COMMENT
 On Dec 16, 1999 this sequence version replaced gi:5763807.
 For more information about this sequence or the Malaria Project,
 see <http://www.sanger.ac.uk/projects/P.falciparum>. IMPORTANT: This
 sequence is unfinished and does not necessarily represent the
 correct of this data. Work on the sequence is in progress and the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc.

FEATURES
 source
 location/Qualifiers
 1..67970
 /organism="Plasmodium falciparum 3D7"
 /strain="3D7"
 /db_xref="taxon:36329"
 /chromosome="1"
 /complement(join(1748..2598,2748..2848,2990..3276))
 /gene="MALIP3.01"
 /complement(join(1748..2598,2748..2848,2990..3276))
 /gene="MALIP3.01"
 /note="MALIP3.01, conserved hypothetical protein, len: 412
 aa, similarity: UP0006 family eg to
 YBLO53C/YBLO512/YBLO511, YBF5, YEA5T (418 aa), fasta
 scores: Opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
 overlap)"
 /codon_start=1
 /product="conserved hypothetical protein, UP0006 family"
 /protein_id="CAB63556.1"
 /db_xref="GI:6594244"
 /translation="MKLVHYKYIVNLFYISLIFKSNLSKLYNDLRISTYVKKY
 LQIKRSNLIKNNIRKMDNESSFLIDGSLNDKMDGVYNSKRHENDLQVNLAK
 NNWVDKLIIFCTCLAEIDKSLKICETDPEGKFLYSAGVNPNCIEFLDKRHEKE
 IIAKKEVEERIKYFKNDQVNSRKNKICDGEKDMNLMLEKRLDIRIPGRKY
 NEKDKYELNKKIKYIPRIYIGIADPEFCSEYDIQIKFIFOLKLVGMEN
 LPMFLHRNCSSETPFKIVDIYKFLPKNGGVHSFDEKEDIYHLYQVKNKLYIGVNG
 CSIKSLKINIAVKRIPLNLLLETPDAPWCVCVKRTMSYEVINDYIKRAYTLMKIKYN
 IICDDNTIKENEPYNTA"
 /complement(2599..2610)
 /gene="MALIP3.01"
 /note="potential splice acceptor sequence"
 /complement(2742..2747)
 /gene="MALIP3.01"
 /note="potential splice donor sequence, atg/gtttaa"
 /complement(2849..2861)
 /gene="MALIP3.01"
 /note="potential splice acceptor sequence"
 /complement(2984..2989)
 /gene="MALIP3.01"
 /note="potential splice donor sequence, aaa/gtataaa"
 5005..5496
 /gene="MALIP3.02"
 5005..5496
 /gene="MALIP3.02"
 /note="MALIP3.02, hypothetical protein, len: 163 aa,
 contains possible signal sequence"
 /codon_start=1

misc_feature
 misc_feature
 misc_feature
 misc_feature
 gene
 CDS

Oy	1531	TTGCGTACAGTTTTAACTTAATTCOCGGTTAGAGATACAAGAAGCATACAGATAACGATT	1530
Db	9259	ATTAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	9318
Oy	1591	AACGGTTACAAAACACTGACGTAAATATATATTTCTCACATTTGTCCATCAGCTCA	1650
Db	9319	ATTATTAGAACATATATATAAATAATATATATATATATATATATATATATATATATATAA	9378
Oy	1651	TGTATTAAGTATGTGTCCTTCCATCACTTAACGGAAGGAATTGACACAAATTTT	1710
Db	9379	ATAAAAACAAATATATATATTTATATATTTTAAATATATATAATTTT-----CATTTT	9431
Oy	1711	ATTATTCGCATCTGATCTAGTTGTATATGAATACCGTGTGAACCCCTTATGACCCC	1770
Db	9432	ATTATATATATTTTATTAATATATATTTTAAATATAATAATTAACCATTTATTAATAT	9491
Oy	1771	AGTTATTCGACCAGCATGCCCTTAGCTAAATGGCATTTGAAACGAAAGATGATPAAG	1830
Db	9492	TATTTATTTTAAATATAACTTTATATATATTTAAATATAATATATATATATATATTTT	9551
Oy	1831	ATGTTGTACGATCATATAGAATATGCAATGATGAGTAAGTAAAGCATTAATATGCT	1890
Db	9552	AAATATATATATAAATATATATATATATATTTTATTAATTAATTAATTAATTAATTA	9611
Oy	1891	TTCTCGTGGTATACGGAAAAAGCCCTTGATPACAGCGTCGATTTTGAACCTTGTAAGT	1950
Db	9612	AATG-TGATATATATATATATATATAT-ATTATATATGTAATATATAATATATATTAAT	9669
Oy	1951	GAACAATCTTGGCGGTGCATTTATGGAACAACATAAAGTANCOTAAATGACCAGGT	2010
Db	9670	TTTTATTTTATATATATATATATATTTTATATATAAATTDATTAAATTAATTAATAATA	9729
Oy	2011	ACAAATATCGATATACGCTATAAAGCTGAANTGTATGTAAGTAATGATGATAT	2070
Db	9730	ATAATATATTAATTAACCTTAATAAT--AATATTTATTTAATATATATATATATCATCAT	9787
Oy	2071	TTTACGTATGAATCTTATATGAAGCTGSAATTTATATTAATAAAGCCCTAGAGAAATTC	2130
Db	9788	TATTTATATCCATTTTATTTTAAATTAATTTATTTATTTTAAATAATATA--TTAATTA	9844
Oy	2131	GACCAATGCAATTTATATGATATCTTTATTTATGTAAGTAAGTACATCGTATTTCACT	2190
Db	9845	ATATATATATAAATATATATTTTATTTTATATATATAATAAATTTATATTTTAATTA	9904
Oy	2191	AACGACTATTT 2202	
Db	9905	ATATACCTATAT 9916	

RESULT	9
AC069525	
LOCUS	
DEFINITION	AC069525 177293 bp DNA linear HIG 24-AUG-2002
ACCESSION	Homo sapiens chromosome 3 clone RP11-306L14, WORKING DRAFT
SEQUENCE	AC069525 2 unordered pieces.
VERSION	AC069525.14 GI:220203149
KEYWORDS	HTG: HTGS_PHASE1; HIGS_DRAFT; HIGS_ACTIVEIN.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 177293)
AUTHORS	Muzny,D.M., Adams,C., Adio-Ogulu,B., Ali-osman,F.R., Allen,C., Alshrocks,S.L., Amaraltinge,H.C., Are,J.R., Ayale,M., Banks,T., Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,U., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Butch,p., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Barhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotte, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frintz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homi, P., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtege, O., Lieu, C., Liu, J., Liu, J., Liu, Y., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Maxhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbad, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okunolu, G., Oranuy, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Plimus, E., Pu, L.L., Quiles, M., Ren, Y., Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, C., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sotaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zortilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished
2 (bases 1 to 177293)
Worley, K.C.
Direct Submission
Submitted (02-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 177293)
Worley, K.C.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Aug 10, 2002 this sequence version replaced gi:20335681.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center Project name: HBMW
Center clone name: RP11-306L14
Summary Statistics
Sequencing vector: M13
Chemistry: Dye-primer Body: 14% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 176749 bases at least Q40
Consensus quality: 176888 bases at least Q30
Consensus quality: 176943 bases at least Q20
Estimated insert size: 168563; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draht_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence as soon as it is available and the accession number will

* be preserved.
1 82664: contig of 82664 bp in length
* 82665 82764: gap of unknown length
* 82765 177293: contig of 94529 bp in length.
FEATURES
source
1. 177293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-306L14"
BASE COUNT 57935 a 37906 c 34820 g 46330 t 102 others
ORIGIN
Query Match 5.8%; Score 132.4; DB 2; Length 177293;
Best Local Similarity 43.6%; Fred. No. 1.6e-09;
Matches 900; Conservative 0; Mismatches 1146; Indels 20; Gaps 6;
QY 130 TATTCGTGACATTTGGTGGTTATGCTTATGCAATTTGATGAGTATTCGAGT 189
DB 80170 TATATTAT 80229
QY 190 TTGCAATCTTTTANGACGCATTTATACGATGTGAAAGACGTGATCAACAA 249
DB 80230 TAT 80286
QY 250 TCAAAACAGATTTATATCATGATTCGACATACGTAATTAATACGATATCTTT 309
DB 80287 AAT 80346
QY 310 CTTAAGACACCATGCGTTATTTGATATATATTTACCGGAAATATATTTAAATCT 369
DB 80347 TAT 80406
QY 370 GCAGAAAGAAAGACACACATTTAAATGAAAGGCGTACTATTTAGCCAAATCA 429
DB 80407 AAT 80466
QY 430 AATGTAATTTGAAAGAAAGCTGTAATCTTCATCAATTTAGATTTAACAATGTTT 489
DB 80467 TAT 80526
QY 490 AATGCTTTAATTTATATTTATGATACCTTTATGAAACCAATACATTCGACAA 549
DB 80527 AAT 80586
QY 550 TAGGTAAAGAAATTTACTGATGGTGTATTTAACAAGATCAACCTTAAACAGATTT 609
DB 80587 AAT 80646
QY 610 AATATACATTTTGTATATAGATATATATATATATATATATATATATATATAT 669
DB 80647 TAT 80706
QY 670 GGAAGATTTTTCGAAACCTGACACCTGACACCTTACATTCATGATGAGCTGAT 729
DB 80707 TATACAT 80766
QY 730 AATATAGATATGTCATATATATATATATATATATATATATATATATATAT 789
DB 80767 AAT 80826
QY 790 TTAACATTTTCATGATATATATATATATATATATATATATATATATATAT 849
DB 80827 TATATACATTTGTATATATATATATATATATATATATATATATATATAT 80886
QY 850 TTAACACCAATATACAGATATTAATATATATATATATATATATATATATATAT 909
DB 80887 ATACAT 80946
QY 910 AATTTTATGACAT 969
DB 80947 AAT 81002

QY 970 TACATTAACAACCAATGATGAGCCTGTTTCTATTGACATCATTCGATTACCT 1029
 DB 81003 TATATTAATAAATAATATTTTATATATATATATATATATATATATATAT 81062
 QY 1030 TATGATATGATGATGATAAATAATCCACATGCTTATCTTTACGACACATATCAGAA 1089
 DB 81063 TATATTTAATAAATATATAAATATATAAATATATATATATATATATATATAT 81122
 QY 1090 CCAAGCTTAACCCATCTTTTATGATGATGATGATGATGATGATGATGAT 1149
 DB 81123 TTTTAT 81182
 QY 1150 AAGATGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1209
 DB 81183 TAT 81242
 QY 1210 ATGTAAATATCTATATATATATATATATATATATATATATATATATATAT 1269
 DB 81243 TAT 81302
 QY 1270 ATGATTCAGACATTCAGGCTATGATTCAGCA-TATACGCTTATATCGTTGTA 1328
 DB 81303 AATTTAT 81362
 QY 1329 ATATGATGCTGATATCTTATGATGACAGGATATATATATATATATATATAT 1388
 DB 81363 AT 81422
 QY 1389 ACAAGTAAAGTAAAAACAATAATGATGATGATGATGATGATGATGATGAT 1448
 DB 81423 AT 81482
 QY 1449 TGATATCAACGACACACCCATCTCAATAGAGGCTTATGATGATGATGATGAT 1508
 DB 81483 AT 81542
 QY 1509 TCGATATATGCAATCTGATGATGATGATGATGATGATGATGATGATGATG 1568
 DB 81543 TAAAT 81602
 QY 1569 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1628
 DB 81603 TAT 81662
 QY 1629 TACATTTGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1688
 DB 81663 TATTTTAT 81722
 QY 1689 AAGTGA--ATGAGACATTTTATATGATGATGATGATGATGATGATGATG 1746
 DB 81723 AT 81782
 QY 1747 GTTGTAAACCCCTTATGATGATGATGATGATGATGATGATGATGATGATG 1806
 DB 81783 AAT 81842
 QY 1807 ATTTGAAACGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1857
 DB 81843 AT 81902
 QY 1858 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1917
 DB 81903 AAT 81962
 QY 1918 GATACAGGCTGATTTGAAACCTTTGATGATGATGATGATGATGATGATG 1977
 DB 81963 TAT 82022
 QY 1978 GAAACCAATA-AAAGTATCTATATGATGATGATGATGATGATGATGATG 2036
 DB 82023 TAT 82082
 QY 2037 TGAATTTGTATGTGATGATGATGATGATGATGATGATGATGATGATGATG 2056

DB 82083 TATATTTTAT 82142
 QY 2097 TGAATTTATATTAAGAAGCGCTAGAGAAATTTGACCATGTCATTTGATGATTC 2156
 DB 82143 AT 82202
 QY 2157 TTAATTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2182
 DB 82203 TTTGAT 82228
 RESULT 10
 AC120883
 LOCUS
 DEFINITION
 Homo sapiens chromosome 18 clone CTD-2146H24 map 18, WORKING DRAFT
 SEQUENCE, 4 ordered pieces.
 AC120883
 AC120883.2 GI:22123189
 HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 133877)
 Birren B., Nusbaum C. and Lander E.
 Homo sapiens chromosome 18, clone CTD-2146H24
 Unpublished
 2 (bases 1 to 133877)
 REFERENCES
 Birren B., Linton L., Nusbaum C., Lander E., Allen N.,
 Anderson S., Bhatia V., Bhatia V., Bloom T., Boudreau N.,
 Boudreau N., Brown A., Cameron A., Campbell A., Chang J.,
 Chazaro B., Chong J., Collins S., Collins S., Collins S.,
 Cook A., Cooke P., DeArnellano K., Dewar K., Diaz J.S., Dodge S.,
 Fero S., Ferreira P., Fitzhugh M., Gage D., Galagan J., Gardyna S.,
 Ginde S., Gord S., Goyette M., Hume W., Ilijev I., Johnson R., Jones C.,
 Hago S., Horton L., Hume W., Ilijev I., Johnson R., Jones C.,
 Karmali A., Karatas A., Kells C., Labocque K., Lamazares R.,
 Landers T., Lebeckzy J., Levine R., Lindblad-Toh K., Liu G.,
 Maclean C., Macdonald P., Major J., Marquis N., Matthews C.,
 McCarty M., McKernan K., Meldrum J., Meneses L.,
 Mihova T., Mlenga V., Murphy T., Naylor T., Nguyen C., Nicol R.,
 Norbu C., Norman C.H., O'Donnell P., O'Neill D.,
 Oliver J., Peterson K., Phunkhang P., Plerre N., Pollara V.,
 Raymond C., Retta M., Riley R., Rize C., Rogov P.,
 Roman J., Rosetti M., Roy A., Santos R., Schauer S., Schupack R.,
 Seaman S., Severy P., Spencer B., Stange-Thomann N., Stojanovic N.,
 Strauss N., Subramanian A., Talamas J., Testaye S., Theodore J.,
 Topham K., Travers M., Travis N., Trifillo J., Vassiliev H.,
 Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye W.J., Young G.,
 Zainoun J., Zembek L., Zimmer A. and Zody M.
 Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 133877)
 REFERENCES
 Birren B., Nusbaum C., Lander E., Allen N., Anderson S.,
 Bhatia V., Bhatia V., Bloom T., Boudreau N., Boudreau N.,
 Brown A., Cameron A., Campbell A., Chang J., Chazaro B.,
 Chong J., Collins S., Collins S., Collins S., Cook A., Cooke P.,
 DeArnellano K., Dewar K., Diaz J.S., Dodge S., Fero S.,
 Ferreira P., Fitzhugh M., Gage D., Galagan J., Gardyna S.,
 Ginde S., Gord S., Goyette M., Hume W., Ilijev I., Johnson R.,
 Jones C., Karmali A., Karatas A., Kells C., Labocque K., Lamazares R.,
 Landers T., Lebeckzy J., Levine R., Lindblad-Toh K., Liu G.,
 Maclean C., Macdonald P., Major J., Marquis N., Matthews C.,
 McCarty M., McKernan K., Meldrum J., Meneses L.,
 Mihova T., Mlenga V., Murphy T., Naylor T., Nguyen C., Nicol R.,
 Norbu C., Norman C.H., O'Donnell P., O'Neill D., Oliver J.,
 Peterson K., Phunkhang P., Plerre N., Pollara V., Raymond C.,
 Retta M., Riley R., Rize C., Rogov P., Roman J., Rosetti M.,
 Roy A., Santos R., Schauer S., Schupack R., Seaman S., Severy P.,
 Spencer B., Stange-Thomann N., Stojanovic N., Strauss N.,
 Subramanian A., Talamas J., Testaye S., Theodore J., Topham K.,
 Travers M., Travis N., Trifillo J., Vassiliev H., Viel R., Vo A.,
 Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J.,
 Zembek L., Zimmer A. and Zody M.

QY 1260 TACATTAAGATGATTCAGACATTCAGGATATGATTCATGATCATGATCATGATTAATTC 1319
 Db 95774 TAAATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 95833
 QY 1320 GTTGTATATATGAATGGAATACCTTCACGACGATGATTAATTTTCAAACTATT 1379
 Db 95834 TTAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 95893
 QY 1380 TATTAAACACGATTAAGTAAAAACAAATATGACATCATCCCTAGACATCA 1439
 Db 95894 TTAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 95953
 QY 1440 CATTACTGATGATTCACGACACACCCATCCTCAATGAGAGATGATGATTAAGT 1499
 Db 95954 ATTATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96010
 QY 1500 CGTTTAATGATTTATAGGACATACCTGATTAATGATTAATTAATTAATTAATTAAT 1559
 Db 96011 AATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96070
 QY 1560 AGATGATTAACAAATGAACATTAACATTAACATTAACATTAACATTAACATTAACATTA 1619
 Db 96071 AATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96130
 QY 1620 ATATATCTGATTTGTCACATCACGTC---ATGTAATACCTTAATGATTCCTTCC 1675
 Db 96131 AATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96190
 QY 1676 AATGATTAACGAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 1735
 Db 96191 AATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96250
 QY 1736 ATATGAATTCGTTGTTAAACCTTATTGACCCGATTTATGACCCGATTAAGCCTTAG 1795
 Db 96251 AATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96310
 QY 1796 GTTAATGGAATTTGTTAAACGAAATGAATGAATGAATGAATGAATGAATGAATGAATGA 1855
 Db 96311 TAAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 96370
 QY 1856 ATGCATATGAAGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 1915
 Db 96371 AATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 96430
 QY 1916 TT 1917
 Db 96431 TT 96432

RESULT 11
 PFMALIP3/c 67970 bp DNA linear INV 15-DEC-1999
 LOCUS Plasmodium falciparum MALIP3, complete sequence.
 DEFINITION AL0311746
 ACCESSION AL0311746 GI:6594243
 VERSION HTG.
 KEYWORDS Plasmodium falciparum 3D7.
 SOURCE Plasmodium falciparum 3D7.
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 67970)
 AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.,
 and Barrett, B.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium,
 The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA, UK

COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807.
 For more information about this sequence or the Malaria Project,
 see http://www.sanger.ac.uk/Projects/P.falciparum. IMPORTANT: This
 sequence is unfinished and does not necessarily represent the
 correct sequence. Work on the sequence is in progress and the
 release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.

FEATURES

source

Location/Qualifiers

1..67970

/organism="Plasmodium falciparum 3D7"

/strain="3D7"

/db_xref="taxon:36329"

/chromosome="1"

gene

complement(join(1748..2598,2748..2848,2990..3276))

/gene="MALIP3.01"

complement(join(1748..2598,2748..2848,2990..3276))

/gene="MALIP3.01"

/note="MALIP3.01, conserved hypothetical protein, len: 412

aa, similarity: UPF0006 family eg to

YB1055C/YB10512/YB10511, YB5_YB5T8 (418 aa), fasta

scores: opt: 316, E(1): 1.1e-12, (33.2% identity in 271 aa

overlap)"

/codon_start=1

/product="conserved hypothetical protein, UPF0006 family"

/protein_id="CAB63556.1"

/db_xref="GI:6594244"

/translation="MKLVPEYIKYIVAVLEYSITPLKSNKIKYNDRIYSVKKY

LDIKRKNLKNKRNIRKMDNESPFDIGSNLTDKMPDGYNNKRNHNDLVNLRK

NNVDKLTITCTCLAEIDSLKICEYDPEKFLYSAGVHPNICEYIDKNHEKE

ILAKKEYEERIKYFKNEQVNSKMGNNKICDGEKDMNMLELKNLDTIPGRKY

NEKDREYLENKKRIKYPNRIVGIEIGIDPRLYFCSKSYIQIKYIFQLKQWEN

LPMLFLHRCNSFEFFKIVDVKFLPEKNGSVISFPKDEIVIIIVQNKNIYGVG

CSLKSLEINAAVKRIPLNLLLETTDAPMGCVKKTASVEYIKDYERKAYTLKIKN

IKCDNTEIKERNEPYNIA"

complement(2599..2610)

/gene="MALIP3.01"

/note="potential splice acceptor sequence"

complement(2742..2747)

/gene="MALIP3.01"

/note="potential splice donor sequence, atg/gttaa"

complement(2849..2861)

/gene="MALIP3.01"

/note="potential splice acceptor sequence"

complement(2984..2989)

/gene="MALIP3.01"

/note="potential splice donor sequence, aaa/gtaaaa"

5005..5496

/gene="MALIP3.02"

/note="MALIP3.02, hypothetical protein, len: 163 aa,

contains possible signal sequence"

/codon_start=1

/product="hypothetical protein, MALIP3.02"

/protein_id="CAB63557.1"

/db_xref="GI:6594245"

/translation="MKLLNRRFVYDPIITLFFPLNSVYLGNNRNNTTFHTEAAK

AMRKLISGEINSLIKDNGDEKIKLNDERKHDSFTKSDKSYFSLNLEEKYSQDLFR

KKQELNEANRKIIEDRQEFYILNNDENIATRFLENNDDELVIOSFQSDIDIIOS

LNN"

8020..10389

/note="possible cent, region of very high [A+T] content"

14884..20352

/gene="MALIP3.03"

14884..20352

/note="MALIP3.03"

/note="MALIP3.03, putative ABC transporter, len: 1822 aa"

/codon_start=1

/product="putative ABC transporter"

/protein_id="CAB63558.1"

/db_xref="GI:6594246"

/translation="MTTYKENVGISKGGKKKKSCONISFLNPLSHDWIRPLINDLIK

GDQELPNCRNDVYVASKLEENRDLEVEDSEYSEKSNSEHYVYKMLHSCNSNASE

KKYVYVYNIIMLSIKTKFRILISYLETILYLGAGFVYKMLHSCNSNASE

YISFLKDFVFSGLVVMTPFHLFEALHLHFPLFTNLKVSIMYFLYKINLCSN

NHONDAFYNTYRKRSQTEIDELSDSLTGKNSSSSGIKNNKKNIDNNKVEN

DYLIINIKSTKKKREKSNLNSNLSPPVNTYNNMESVPSYTFYSCILFVWPYKIF

MSYVYHRIKIGSNVGIATWLSIALTSLATLEFPLSPKSKYLLIYRDRIDNNHVL

D	b	24296	TAAATATTTTTAACTGTAATAAATTAATAATA---TTTAACTAATTAATAAT	24240
Q	y	1302	GCAATACGGTAAATCGTTGTATATAGAAATGAATACTTCATGCAGGTAAAT	1361
D	b	24239	AAAATATTATTAACGTGATAAATATAAATATATTATATTATAACGTAAAAATATA	24180
Q	y	1362	TATTTTCCAACATATTTTATTTAAACACAGGTAACTTAAACCAAATCATATGA-	1420
D	b	24179	ATTATTAAATATTAATATAAATATAAATATAAATTTATTAATATTTAAAAATATTA	24120
Q	y	1421	CATCACCTTAGCAGCTATCATTAACGATATCAACGACGCCATCTCAATGAG	1480
D	b	24119	TATTAATATATTAATATAAATATAAATATAAATATAAATATAAATATAAATA	24060
Q	y	1481	AGGTAACTTATCTAAAGCTTTTAATGATTATAGCATACCTGCAT-TACGTTCA	1539
D	b	24059	TATTAATATAATATAAATATAAATATAAATATAAATATAAATATAAATATAA	24000
Q	y	1540	CATTTTAACTTATCCGTTAGATATCAACATGAATCAATCAATATCATTAACGTTAC	1599
D	b	23999	CACATAATATATTAATATAATATAATATCTATATTTAAATATTAATATAATATAT	23940
Q	y	1600	AAAAACAGTAAGTAATATATATTCTCTACATTGTGCACATCAGCTGATTGAAC	1659
D	b	23939	AGTATATATAATATAATATAAATATAAATATAATATAATATAATATAATATAAT	23880
Q	y	1660	TTATGTCTCTTCCAAATACCTTAACGAAAGTGAAATGACGCAATTTATTTTTC	1719
D	b	23679	TAAATATTAATATAAATATAAGTTTATATATAATTAATATAATATATATTAGAATA	23620
Q	y	1720	GATACGTATAGTTGTATATGAATCCGTTGTAAACCCTTATGAAACCCCATTTATTC	1779
D	b	23819	TTCAATATAATATAATATAATATAAATATAATATAATATAATATAATATAATAT	23760
Q	y	1780	GACCCGATAGCCTAGCTAAAGGATATGAAACACGACAGATAGATAAGATGTTTGA	1839
D	b	23759	ATAATTTTAAATATAAATAATATAAATTAATATAATATAAATATAAAT-ATAAGTATA	23701
Q	y	1840	CTGATCATAGAAATATCATATGAAGTGAATGGAAGATTAAATATCGTCTCGTST	1899
D	b	23700	TTAAATATAATATAATATATATTTTAAATATAAATATAAATATAAATATAAATA	23641
Q	y	1900	ATACGAAAAAGCCCTTGATACAAGCTCAT-----TTGAAACCTTTGACGTGA	1953
D	b	23640	ATAATATATGATATATAATATAAATATAATATAATATAATATAATATAATATA	23581
Q	y	1954	CAATTCCTTGACGCGCCATTTATGAAACATPAAAGTATATCTATATGSCAAGTACA	2013
D	b	23580	ATATATTTTAAATATAAATATTTTAAATATAATATAATATAAATATAAATATA	23521
Q	y	2014	ATAATGATATATCCGTATAAACCTGAATGTATATGCTGAATGATATGATATTTT	2073
D	b	23520	ATAATATAATATAATATAATATAATATAATATAAATATAAATATAAATATAA	23462
Q	y	2074	ACGATAGAACTTAATATGAACGTAATTTATTTAAAGCCGTAGAGAAATTTCCGC	2133
D	b	23461	-----ATATTAATATAAATATATGTTTATAATATAATATAATTTAAATATAAATAT	23409
Q	y	2134	CATAGCTCAATTTGATGATATTCCTTATATTAAGATGACATCGGTCATTTCACTTAAC	2193
D	b	23408	ATTAAATATAAATATATGTTTATATAATATATTTAAATATAAATATAAATATA	23349
Q	y	2194	GACTTATTTCCAGTGAAGCTCAGTACATACAAAATCTGATTTGATATATTAAGCT	2253
D	b	23348	ATAATATTTTATAATAAATAATATAATATAATATAATATGTTTATATAAATATAT	23289
Q	y	2254	GAACATGATGAAT 2267	
D	b	23288	AAATATAAATATAT 23275	

REFERENCE 5 (bases 1 to 181792)
 AUTHORS Waterston,R
 TITLE Direct Submission
 JOURNAL Submitted (22-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Feb 1, 2002 this sequence version replaced gi:17977471.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: MUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center Project name: H_NH0338B23
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frenken,E., Tatenio,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-631121, 2000 bp overlap; the clone sequenced to the right is RP11-660813, 2000 bp overlap. Actual start of this clone is at base position 161305 of RP11-631121; actual end is at base position 181792 of RP11-338B23.

There is a region covered by PCR only from 20829 to 20858. There is a simple-sequence repeat from 81554 to 85256.

FEATURES
 source The sequence of AC025095 has been incorporated into AC098822.

Location/Qualifiers
 1. 181792
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-338B23"
 /clone_lib="RPCT-11"
 737. 816
 /rpt_family="MIR"
 repeat_region
 1478. 1769
 /rpt_family="L2"
 repeat_region
 2124. 2250
 /rpt_family="MIR"
 repeat_region
 3056. 3349
 /rpt_family="L1"
 repeat_region
 3185. 3326
 /rpt_family="AT_rich"

repeat_region 3350. 3763
 /rpt_family="L1"
 repeat_region 3743. 3782
 /rpt_family="AT_rich"
 repeat_region 3764. 3790
 /rpt_family="L1"
 repeat_region 3853. 3956
 /rpt_family="L1"
 repeat_region 3971. 4044
 /rpt_family="L1"
 repeat_region 4009. 4031
 /rpt_family="AT_rich"
 repeat_region 4056. 4130
 /rpt_family="ERVVL"
 repeat_region 4276. 4349
 /rpt_family="ERVVL"
 repeat_region 4353. 4908
 /rpt_family="L1"
 repeat_region 5076. 5736
 /rpt_family="L1"
 repeat_region 5737. 6109
 /rpt_family="MALR"
 repeat_region 6110. 6569
 /rpt_family="L1"
 repeat_region 6594. 6907
 /rpt_family="Alu"
 repeat_region 6876. 6907
 /rpt_family="(A)n"
 repeat_region 6910. 7076
 /rpt_family="L1"
 repeat_region 7077. 7376
 /rpt_family="Alu"
 repeat_region 7377. 8198
 /rpt_family="L1"
 repeat_region 8201. 8238
 /rpt_family="AT_rich"
 misc.feature 8522. 8729
 /note="match to EST B1005753 (NTD:914409627)"
 repeat_region 8701. 8813
 /rpt_family="MERL_type?"
 repeat_region 8899. 9178
 /rpt_family="MERL_type?"
 repeat_region 9264. 9415
 /rpt_family="CT_rich"
 repeat_region 9408. 9713
 /rpt_family="Alu"
 repeat_region 9570. 9594
 /rpt_family="(T)n"
 repeat_region 9718. 10701
 /rpt_family="L1"
 misc.feature 10384. 10877
 /note="match to EST A1243631 (NTD:93839028) qh88a07.x1"
 repeat_region 10918. 10952
 /rpt_family="AT_rich"
 repeat_region 11758. 11794
 /rpt_family="AT_rich"
 repeat_region 11940. 12151
 /rpt_family="MIR"
 repeat_region 12171. 12195
 /rpt_family="(TAA)n"
 repeat_region 12391. 12419
 /rpt_family="AT_rich"
 repeat_region 13031. 13175
 /rpt_family="L2"
 repeat_region 13761. 14012
 /rpt_family="MALR"
 repeat_region 14013. 14050
 /rpt_family="ERVVL"
 repeat_region 14051. 14075
 /rpt_family="(T)n"
 repeat_region 14361. 14556
 /rpt_family="ERVVL"
 repeat_region 14546. 14662

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeKrellano, K., Dewar, K., Diaz, J., S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, N., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Willson, B., Wu, X., Wyman, D., Ye, W., J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (pages 1 to 176898)

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (pages 1 to 176898)

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeKrellano, K., Dewar, K., Diaz, J., S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, N., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Willson, B., Wu, X., Wyman, D., Ye, W., J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
 Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 1, 2002 this sequence version replaced g1:20369476.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L22018
 Center clone name: 784_B_15

Only the last 176.9 kilobases of this clone are being submitted.
 The remainder overlaps accession number AC091321 [WICR project L13163].

FEATURES
 source

Location/Qualifiers

1..176898
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-784B15"
 /clone_11b="RP11-11 Human Male BAC"
 27..70
 /rpt_family="AT_rich"
 528..682
 /rpt_family="LIMB5"
 683..711
 /rpt_family="LIMB5"
 712..1048
 /rpt_family="LIMB5"
 complement(1365..1726)
 /rpt_family="L1"
 1770..1906
 /rpt_family="MER69A"
 complement(2633..2985)
 /rpt_family="LIME"
 complement(3047..3308)
 /rpt_family="LIME"
 3734..3847
 /rpt_family="MLT1D"
 complement(3848..4056)
 /rpt_family="MS1D"
 complement(4110..4183)
 /rpt_family="MS1D"
 4184..4532
 /rpt_family="MLT1D"
 complement(4799..4918)
 /rpt_family="L1"
 complement(5584..6121)
 /rpt_family="ERV1"
 complement(6193..6578)
 /rpt_family="MLT1"
 complement(6621..6705)
 /rpt_family="MLT1-int"
 6969..7174
 /rpt_family="Alu10"
 7575..7599
 /rpt_family="AT_rich"
 8207..8326
 /rpt_family="(TA)n"
 8334..8357
 /rpt_family="AT_rich"
 8366..8393
 /rpt_family="(TC)n"
 complement(8432..8791)
 /rpt_family="MLT1"
 complement(8806..8839)
 repeat_region

Query Match	5.4%;	Score 122.6;	DB 9;	Length 176898;
Best Local Similarity	43.3%;	Pred. No. 3.2e-08;		
Matches 831;	Conservative	0;	Mismatches 1079;	Indels 11;
				Gaps 5

Db	11811	ATATTTTATTAATAAGAAATATATATCTGTATAAATATATAATATATAAATATTAATATATAT	11870
Qy	679	TTTGCAAAGCTCACACCTGGAACAACCTTACATACATTCATATATGACGTGATTAATTAGGT	738
Db	11871	TATATTAATATATAATATAATAAAAAATATATATATATATTTATATAAAATATATATGTAATAATA	11930
Qy	739	ATGTCCCATTTTCATTAATAGTATATATTTTCCAAATTTTGACCTTAAACCAATTAACGTTT	798
Db	11931	TAAATATATATATATAATATAATATATATAAAAAATATATATATATATTCATATATAATACATA	11990
Qy	799	TCATGAATATATTATGGAATCTTACTGAAATATGAAATGACACGTTTTCAGTTACTCAC	858
Db	11991	TATATAAATATAAATATATATATATATATAAATATAAATATATAATATATATATATATATATA	12050
Qy	859	CAATTCGAAGTAT-TAAATATCTTATACCATTAATCATTTCCATCATATGGAATTTTGA	917
Db	12051	TAAATATATTTATATATATATATATATATATATATAAATATATAAATATATATATATATATA	12110
Qy	918	TCAGTATATTTAAACATTCATCTCGTGGTGTATTAATATGATATACACCAACATACATAAA	977
Db	12111	TAAATATATAAATATATATATATATATATAATATAATATAAATATATATATATATATATATAT	12170
Qy	978	CAAACTAATTCATGACGCTGTTTCTATATTCATGACATCAATTCGAGTTATGCTTATGAT	1037
Db	12171	GAAATATATTTATATATAAATATATATATATATATATATATATATGAAAATATATTATATAA	11230
Qy	1038	GTATCATCAAAAATATCCACATCGTTATACCTTTTACCAACAGTATCAACACCAACGTT	1097
Db	11231	ATAAATTTATATTTATATATATATATAATATATATAATATATAATATAAATATATATTTAT	11290
Qy	1098	AATCCCTACCTTTTATGATGATGACAAATTATTTTTCATATATAGATGATGAAGATGT	1157
Db	11291	ATAATATATAAATATATATATATATAAATATAAATATATATATATAATATATAATATAATAT	11350
Qy	1158	ATTTAAGCATATTTATTAATTTAAATTTAAATCCGCTGTTCGCGCAATGATGTAATAA	1217
Db	112351	ATTATATTTTATATATATATATATAAATATATATATATATTTATTTATATAAATATAAATATA	112410
Qy	1218	ATACATATATATGATATATGATTCGTAAATATCATCAAAATACATTAAGAATGATTCAT	1277
Db	112411	TATATATATTTATATATATATATATATGTTATATATAAATATATATATATATTTATATG	112470
Qy	1278	AGACATTCAGGCT--ATGATTCGATGCAATATACGCTGTAATTCGTTGTTTATATATGA	1334
Db	112471	ATATATTTATATATATAATATATAATATATAATGAAATATAATATATATATATATATTTACTA	112530
Qy	1335	ATGCAATACCTTCATGACGACGTGATATTTTTCAAACGATTTTATTAACACACAG	1394
Db	112531	ATATATATATATATATATATATATATATATATAATAATACATATATAATGCAATATATAA	112590
Qy	1395	TAACTTA-AAAACAAAATCANATGACATCACCTTACGACATACATTACTGATGATTA	1453
Db	112591	TATATTCATATAAATACAT	112650
Qy	1454	TCAAGCAACACCCATGACCAAGGAGAGGTATGTTATCTGAAAGCGTTTAAATGGAAT	1513
Db	112651	AAATACATATATATATATATATATATAAATACAAATATATAATATATATATAATACATA	112710
Qy	1514	TATATGGCATCCGCAATTACGTTACATTTTAACTTATTCGTTAGATGATACAGATG	1573
Db	112711	TATATTTATATAAATATATATATATATATATTTTATAAATAATATATATATATATATAA	112770
Qy	1574	AACATATACAAATATTAACGGTTACAAAAACACGTGAAGCTAATATATTTCTCTACAT	1633
Db	112771	AATATATATATATTTTATATAAAATATATATTTTATATATAAATAATATATATTTATAT	112830
Qy	1634	TTTGACATACGCTTCATTTGATATACCTATATGGTTCCTTCCAACTACTTAACGGAAGTG	1693
Db	112831	ATATATATATCATATATATATATAATATATATATATATGCT-----ATATATATATAAGATA	112885
Qy	1694	AAATTCAGACATTTTATTTATTTATTCGATACCTGATAGTTTGTATATGAAATCCGTTGTTA	1753

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2003, 23:54:27 ; Search time 314 Seconds
(without alignments)

16395.122 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 2286
Sequence: 1 atggattactagatgcatt.....taaaaaaggaactgttaa 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

```

1: N.Geneseg_101002:*
2: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseg/geneseg-n-emb1/NA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2286	100.0	2286	21	AAA69014	Bacteriophage 44AH
2	2286	100.0	16668	21	AAA69013	Bacteriophage 44AH
3	222	9.7	222	21	AAA69032	Bacteriophage 44AH
4	153	6.7	153	21	AAA69048	Bacteriophage 44AH
5	129	5.6	129	21	AAA69065	Bacteriophage 44AH
6	129	5.6	129	21	AAA69066	Bacteriophage 44AH
7	123	5.4	123	21	AAA69070	Bacteriophage 44AH
8	111	4.9	111	21	AAA69077	Bacteriophage 44AH
9	102	4.5	102	21	AAA69086	Bacteriophage 44AH

C 10	98.6	4.3	5930	24	ABL32517	Human immune syste
C 11	95.4	4.2	7597	24	ABL33013	Human immune syste
C 12	95.4	4.2	32392	24	ABL56203	AMEPV genome fragm
C 13	94	4.1	11691	24	ABL34240	Human immune syste
C 14	93	4.1	15548	24	ABL34155	Human immune syste
C 15	93	4.1	50000	24	ABL35644	Human immune syste
C 16	90.2	3.9	50000	24	ABL55643	AMEPV genome fragm
C 17	89.8	3.9	32392	24	ABL56203	AMEPV genome fragm
C 18	89.6	3.9	6071	24	ABL32325	Human immune syste
C 19	89.6	3.9	6071	24	ABL561076	Human gene regulat
C 20	87.6	3.8	6013	24	ABL561265	Human gene regulat
C 21	87.6	3.8	6013	24	ABL33361	Signal transductio
C 22	86.6	3.8	12069	24	ABL39930	Human chemically p
C 23	86.6	3.8	73334	24	ABL92319	Chemically treated
C 24	86.6	3.8	73334	24	ABL34125	Human immune syste
C 25	86.4	3.8	1830	24	ABL56243	AMEPV metalloprote
C 26	85.8	3.8	11922	21	AAA70187	Plasmodium falcipa
C 27	85.4	3.7	20420	22	AAK73165	Human immune/haema
C 28	85.4	3.7	20420	22	ABK69933	Human secreted pro
C 29	85.2	3.7	7892	24	ABK40056	Human chemically p
C 30	85.2	3.7	11812	22	AAK45502	Chemically pretrea
C 31	85.2	3.7	11812	22	AAK46742	Tumour suppressor
C 32	85.2	3.7	11812	24	ABL34119	Human immune syste
C 33	85.2	3.7	11812	24	ABK28432	DNA transcription
C 34	85	3.7	18154	24	ABL32254	Human immune syste
C 35	85	3.7	19124	18	AAK72882	Human immune syste
C 36	85	3.7	19124	21	AAK98287	Plasmodium var-7 g
C 37	84.8	3.7	5413	22	AAK46694	Tumour suppressor
C 38	84.8	3.7	6092	24	AAK61080	Human gene regulat
C 39	84.8	3.7	47108	24	ABK31511	Signal transductio
C 40	84.6	3.7	50000	24	ABL56202	AMEPV genome fragm
C 41	84	3.7	6033	21	AAA70152	Plasmodium falcipa
C 42	84	3.7	7814	22	AAK46530	Tumour suppressor
C 43	83.8	3.7	15387	24	ABL32184	Human immune syste
C 44	83.2	3.6	6175	24	ABL33307	Human immune syste
C 45	82.2	3.6	7455	24	ABL33758	Human immune syste

ALIGNMENTS

```

RESULT 1
ID AAA69014 standard; DNA; 2286 BP.
XX AAA69014;
AC AAA69014;
DT 27-OCT-2000 (first entry)
XX
DE Bacteriophage 44AHJD nucleotide sequence 44HJDF001.
XX
DE Bacteriophage 44AHJD nucleotide sequence 44HJDF001.
XX
KM Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; bacterial infection; ds.
XX
OS Bacteriophage 44AHJD.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PE 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157216.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M.
XX

```

DR WPI: 2000-412361/35.
 DR P-PSDB: AAB16529.
 XX Identifying a bacteriophage coding region for treating bacterial
 PT infections comprises identifying a nucleic acid encoding a product that
 PT inhibits bacteria when a bacteriophage infects a bacterium
 XX
 PS Example 9; Page 272; 456pp; English.

CC The present invention describes a method for identifying a bacteriophage
 CC coding region encoding a product active on an essential bacterial
 CC target. The method comprises identifying a nucleic acid sequence encoding
 CC a gene product that provides a bacteria-inhibiting function when an
 CC uncharacterised bacteriophage infects a pathogenic bacterium. The
 CC compound active on a target of a bacteriophage inhibitor protein in a
 CC bacteria is used to treat or prevent a bacterial infection in an animal.
 CC AAB68243 to AAB69442 and AAB16523 to AAB16954 represent bacteriophage
 CC nucleotide and protein sequences which are used in the exemplification of
 CC the present invention.

SQ Sequence 2286 BP; 866 A; 315 C; 323 G; 782 T; 0 other:

Query Match 100.0%; Score 2286; DB 21; Length 2286;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGATTCGTGAATGCAATGCAATATCAATGAACGCGAATGTTTATACGG 60
 Db 1 ATGGGATTCGTGAATGCAATGCAATATCAATGAACGCGAATGTTTATACGG 60
 QY 61 GATATGAAACATTTAGCGTACATTAAGTTAAACGCAAAAAAACAACCAATATATA 120
 Db 61 GATATGAAACATTTAGCGTACATTAAGTTAAACGCAAAAAAACAACCAATATATA 120
 QY 121 AAGGTACTATTCGTACCAATGCTGTTAAGTTAAGTAAATGATGATTTGAAGA 180
 Db 121 AAGGTACTATTCGTACCAATGCTGTTAAGTTAAGTAAATGATGATTTGAAGA 180
 QY 121 AAGGTACTATTCGTACCAATGCTGTTAAGTTAAGTAAATGATGATTTGAAGA 180
 Db 121 AAGGTACTATTCGTACCAATGCTGTTAAGTTAAGTAAATGATGATTTGAAGA 180
 QY 181 TTCCGAGTTTGGAACTCTTTTATGACGCAATTTATACGTATGTGAAAAGAGCTATCA 240
 Db 181 TTCCGAGTTTGGAACTCTTTTATGACGCAATTTATACGTATGTGAAAAGAGCTATCA 240
 QY 241 ATGCAAAAATCAAAAACACATTTATCAGATTGCACATACCTGTAATTAATACGATAT 300
 Db 241 ATGCAAAAATCAAAAACACATTTATCAGATTGCACATACCTGTAATTAATACGATAT 300
 QY 301 CATTTTACTTAAAGACACCAATGCTTTTGTATATATATACACGCAAAATATATAT 360
 Db 301 CATTTTACTTAAAGACACCAATGCTTTTGTATATATATACACGCAAAATATATAT 360
 QY 361 TTAANAATCGCAGAGAAAATGACACACATTAAAAAGAGGCTATCTATTTAGCC 420
 Db 361 TTAANAATCGCAGAGAAAATGACACACATTAAAAAGAGGCTATCTATTTAGCC 420
 QY 421 AAAAATCAAAATGTATTTAGAAAACGTTAAATCTTCATCAATTTATATTTACA 480
 Db 421 AAAAATCAAAATGTATTTAGAAAACGTTAAATCTTCATCAATTTATATTTACA 480
 QY 481 ATGTTTTAAATGTTTTAAATTAATTAATGATTAATGATTAATGATTAATGATTAAT 540
 Db 481 ATGTTTTAAATGTTTTAAATTAATTAATGATTAATGATTAATGATTAATGATTAAT 540
 QY 541 ATGCAACATTTAGTAAATTTACTGATGAGGTTATTTAAACGAATCAACACTTAA 600
 Db 541 ATGCAACATTTAGTAAATTTACTGATGAGGTTATTTAAACGAATCAACACTTAA 600
 QY 601 ACAGATTTTAATTAATGATTTTGTATAAGATAATGATTAATGATTAATGATTAAT 660
 Db 601 ACAGATTTTAATTAATGATTTTGTATAAGATAATGATTAATGATTAATGATTAAT 660
 QY 661 GACTATGCTGTGAATGTTTGCAAAACCTCAACCACTTACATACATTTATATAT 720
 Db 661 GACTATGCTGTGAATGTTTGCAAAACCTCAACCACTTACATACATTTATATAT 720

QY 721 GACGTGATTAATTAAGTATGTCATTCATTAATAGATATATTTCCAAATTTTGAC 780
 Db 721 GACGTGATTAATTAAGTATGTCATTCATTAATAGATATATTTCCAAATTTTGAC 780
 QY 781 TATAAACAATTAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
 Db 781 TATAAACAATTAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
 QY 841 CGTTTTCAGTTTACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
 Db 841 CGTTTTCAGTTTACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
 QY 901 CATGATATGAATTTTATGATATATTAATTAATTAATTAATTAATTAATTAATTAAT 960
 Db 901 CATGATATGAATTTTATGATATATTAATTAATTAATTAATTAATTAATTAATTAAT 960
 QY 961 AACACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
 Db 961 AACACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
 QY 1021 AGTTATCCTTATGATGATATCATGAAAAAATTCACCAATGCTTATCTTTACGAAC 1080
 Db 1021 AGTTATCCTTATGATGATATCATGAAAAAATTCACCAATGCTTATCTTTACGAAC 1080
 QY 1081 TATTCAGAACCAAGTTAATCCCTACTTTTATGATGATGACATTAATTTTCATTTAT 1140
 Db 1081 TATTCAGAACCAAGTTAATCCCTACTTTTATGATGATGACATTAATTTTCATTTAT 1140
 QY 1141 AAGATTAAGTAAGTATTTTAACGATATTTAATTAATTAATTAATTAATTAATTAAT 1200
 Db 1141 AAGATTAAGTAAGTATTTTAACGATATTTAATTAATTAATTAATTAATTAATTAAT 1200
 QY 1201 CGTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260
 Db 1201 CGTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260
 QY 1261 ACATTAAGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320
 Db 1261 ACATTAAGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320
 QY 1321 TTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
 Db 1321 TTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
 QY 1381 ATTAANAACCAAGTAAGTTAAAAAACAATTAATTAATTAATTAATTAATTAATTAAT 1440
 Db 1381 ATTAANAACCAAGTAAGTTAAAAAACAATTAATTAATTAATTAATTAATTAATTAAT 1440
 QY 1441 ATTACTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
 Db 1441 ATTACTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
 QY 1501 GTTTTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
 Db 1501 GTTTTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
 QY 1561 GATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
 Db 1561 GATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
 QY 1621 TTATTCCTTACATTTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
 Db 1621 TTATTCCTTACATTTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
 QY 1681 TTAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
 Db 1681 TTAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
 QY 1741 AAATCCGTTGTTAAACCTTTTGAACCCGTTTATTCGACCGGATAGCCTTAGATAA 1800
 Db 1741 AAATCCGTTGTTAAACCTTTTGAACCCGTTTATTCGACCGGATAGCCTTAGATAA 1800

```
QY 1801 TGGGATATTGAAAGCAAGATAGATAGATGTTGTACTGATCATAGAAATATGCA 1860
    |||||||
Db 1801 TGGGATATTGAAAGCAAGATAGATAGATGTTGTACTGATCATAGAAATATGCA 1860
QY 1861 TATGAGTGAATGGAAGATTTAAATTCCTCTGCTGATATACCGAAAAAGCCTTTGAT 1920
    |||||||
Db 1861 TATGAGTGAATGGAAGATTTAAATTCCTCTGCTGATATACCGAAAAAGCCTTTGAT 1920
QY 1921 ACAAGCCTGATTTTGAACCTTGTACGACAAATCTTTGACGCTGGCATTATTGAA 1980
    |||||||
Db 1921 ACAAGCCTGATTTTGAACCTTGTACGACAAATCTTTGACGCTGGCATTATTGAA 1980
QY 1981 AACAAATAAAGTATCTAATATGAGCAAGTACAAATATGATATATCCGTCTAAACTGAA 2040
    |||||||
Db 1981 AACAAATAAAGTATCTAATATGAGCAAGTACAAATATGATATATCCGTCTAAACTGAA 2040
QY 2041 ATTGTATGCTGATATGATATGATGATATTTTACTGATACCTTAATATGAAACGTGAA 2100
    |||||||
Db 2041 ATTGTATGCTGATATGATATGATGATATTTTACTGATACCTTAATATGAAACGTGAA 2100
QY 2101 TTTTATTTAAAGACGCTAGCAAAATTTGACCATAGTCATTAATATGAAACGTGAA 2160
    |||||||
Db 2101 TTTTATTTAAAGACGCTAGCAAAATTTGACCATAGTCATTAATATGAAACGTGAA 2160
QY 2161 ATTGAAGTACATCGGTTTCATTTCACCTTACGACCTTATTTCCAGTTGAACGTTAGTA 2220
    |||||||
Db 2161 ATTGAAGTACATCGGTTTCATTTCACCTTACGACCTTATTTCCAGTTGAACGTTAGTA 2220
QY 2221 CATACAAATCTGATTTGCAATATTTAAACGGAACATGATGAAATAAAGGCAAC 2280
    |||||||
Db 2221 CATACAAATCTGATTTGCAATATTTAAACGGAACATGATGAAATAAAGGCAAC 2280
QY 2281 TGTATA 2286
    |||||||
Db 2281 TGTATA 2286
    |||||||

RESULT 2
AAA69013/c
ID AAA69013 standard; DNA: 16668 BP.
XX
AC AAA69013;
XX
DT 27-OCT-2000 (first entry)
XX
DE Bacteriophage 44ABD complete genome sequence.
XX
KM Bacteriophage; antimicrobial; genome; identification; antibacterial;
XX
OS Bacteriophage 44ABD.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
FA (PHAG-) PHAGEBTECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX
WP: 2000-412361/35.
XX
DR
XX
PT Identifying a bacteriophage coding region for treating bacterial
    infections comprises identifying a nucleic acid encoding a product that
```

```
PT Inhibits bacteria when a bacteriophage infects a bacterium
XX
XX Example 9; Page 266-269; 456pp; English.
PS
XX The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SO Sequence 16668 BP; 6095 A; 2338 C; 2608 G; 5627 T; 0 other;

Query Match 100.0%; Score 2286; DB 21; Length 16668;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGATTTACTAGAAATGCAATATCATTAACATGACAGCTGATGTTTATATCTGG 60
    |||||||
Db 12627 ATGGGATTTACTAGAAATGCAATATCATTAACATGACAGCTGATGTTTATATCTGG 12568
QY 61 GATATGAAACATTAGCGTACATTAAGTTAAACGAGCAAAACCAACCAATATATA 120
    |||||||
Db 12567 GATATGAAACATTAGCGTACATTAAGTTAAACGAGCAAAACCAACCAATATATA 12508
QY 121 AACGTTACTATCTCTAGCAATGTTGTTTAAGTTTGAATATGATGTTGAAGTA 180
    |||||||
Db 12507 AACGTTACTATCTCTAGCAATGTTGTTTAAGTTTGAATATGATGTTGAAGTA 12448
QY 181 TTTCGAGTTTCCGATCTTTTATGACGCAATTTATACGTATGTAAGACGTGATACA 240
    |||||||
Db 12447 TTTCGAGTTTCCGATCTTTTATGACGCAATTTATACGTATGTAAGACGTGATACA 12388
QY 241 ATACCAAAATCAAAAACAGATATTATCATGATTGCAATACCTGATTAATATATAT 300
    |||||||
Db 12387 ATACCAAAATCAAAAACAGATATTATCATGATTGCAATACCTGATTAATATATAT 12328
QY 301 CATTTTCTTAAAGACACACATCGGTTATTTGATATATTTACAGCCGCAAAATATATAT 360
    |||||||
Db 12327 CATTTTCTTAAAGACACACATCGGTTATTTGATATATTTACAGCCGCAAAATATATAT 12268
QY 361 TTTAAATCTGCAAGAAATGACACACATTTAAATGAAGAGGCTACTATTTAGCC 420
    |||||||
Db 12267 TTTAAATCTGCAAGAAATGACACACATTTAAATGAAGAGGCTACTATTTAGCC 12208
QY 421 AAAAATCAAAATGATTTTGAAGAAACGTTAAATCTTCATCATCAATTTAGATTAA 480
    |||||||
Db 12207 AAAAATCAAAATGATTTTGAAGAAACGTTAAATCTTCATCAATTTAGATTAA 12148
QY 481 ATGTTTAAATGTTTAAATTTAATATATGATGATGATTTTGAAGAAACCAATACATCA 540
    |||||||
Db 12147 ATGTTTAAATGTTTAAATTTAATATATGATGATGATTTTGAAGAAACCAATACATCA 12088
QY 541 ATTGCAACATTAGGTAAAGAAATTTGATGTTGTTATTTTAAACAAATCACACTTAA 600
    |||||||
Db 12087 ATTGCAACATTAGGTAAAGAAATTTGATGTTGTTATTTTAAACAAATCACACTTAA 12028
QY 601 ACAGATTTTATATATGATTTTGAATGAAGATATGATATGATGATGATGATGATGAT 660
    |||||||
Db 12027 ACAGATTTTATATATGATTTTGAATGAAGATATGATATGATGATGATGATGATGAT 11968
QY 661 GACTATGCTGTGAATGTTTSCAAAACCTGACACCTGACCACTTACATATTCATAT 720
    |||||||
Db 11967 GACTATGCTGTGAATGTTTSCAAAACCTGACACCTGACCACTTACATATTCATATAT 11908
QY 721 GACGTGATTTATTTAGGTATGTCATATTCATATGATGATGATGATGATGATGATGAT 780
    |||||||
Db 11907 GACGTGATTTATTTAGGTATGTCATATTCATATGATGATGATGATGATGATGATGAT 11848
```

OY	781	TATACAAATTCACTTTCATCGTAAGATATATGSAATCCTCCCTGGAAATAAGGAAGACA	840
Db	11847	TATACAAATTCACTTTCATCGTAAGATATATGSAATCCTCCCTGGAAATAAGGAAGACA	11786
OY	841	CGTTTGAGTTACTGCACCAATATCAAGATATTTAAAATATCTTATACCATTTATCATTTTC	900
Db	11787	CGTTTGAGTTACTGCACCAATATCAAGATATTTAAAATATCTTATACCATTTATCATTTTC	11722
OY	901	CATGATATGAAATTTTTATGACATATATTAATATCATATCTATCGTGGGTGTTTAAATATGTAT	960
Db	11727	CATGATATGAAATTTTTATGACATATATTAATATCATATCTATCGTGGGTGTTTAAATATGTAT	11668
OY	961	AACACCAAATACATAAACAACAACTAATGATGAGCCCTGTTTTTTCTATTTGACATCAATATCG	1020
Db	11667	AACACCAAATACATAAACAACAACTAATGATGAGCCCTGTTTTTTCTATTTGACATCAATATCG	11608
OY	1021	AGTATTCCTTATGATGATATCATGAAAAATTCACAACTGGTATATCTTTTACGACATC	1080
Db	11607	AGTATTCCTTATGATGATATCATGAAAAATTCACAACTGGTATATCTTTTACGACACATC	11548
OY	1081	TATTCAGAACCAACGATTAATCCCTACTTTTTTAGATGATGACAAATTTATTTTACATATAT	1140
Db	11547	TATTCAGAACCAACGATTAATCCCTACTTTTTTAGATGATGACAAATTTATTTTACATATAT	11488
OY	1141	AAGATTGATTAAGATGATTTTAAAGATGATTTTATTAATTAATTAATACGCTGATATTA	1200
Db	11487	AAGATTGATTAAGATGATTTTAAAGATGATTTTATTAATTAATTAATACGCTGATATTA	11428
OY	1201	CGTAAATGATTTAAATATCTATATATATGATATGATATGATTCGTTAATATATCAATACAAAT	1260
Db	11427	CGTAAATGATTTAAATATCTATATATATGATATGATATGATTCGTTAATATATCAATACAAAT	11368
OY	1261	ACATTAAGAATGATTCAGACATTCACGGGATTTGATTCATCATCATATATACGTGTTAATTCG	1320
Db	11367	ACATTAAGAATGATTCAGACATTCACGGGATTTGATTCATCATCATATATACGTGTTAATTCG	11308
OY	1321	TTTGCTTATATGAATATGGAATATCTTTCATGACGTGATATATTTTTCCAAACATATTTT	1380
Db	11307	TTTGCTTATATGAATATGGAATATCTTTCATGACGTGATATATTTTTCCAAACATATTTT	11248
OY	1381	ATTAAAAACAGAGTATGTTAAAAAACAATAATCATATATGACATACCTTCAGCATATCAC	1440
Db	11247	ATTAAAAACAGAGTATGTTAAAAAACAATAATCATATATGACATACCTTCAGCATATCAC	11188
OY	1441	ATTATCTATGATATTCACAGACATCCCATATCTCAANTAGGAGGTTATGTTATCTTAAAGTC	1500
Db	11187	ATTATCTATGATATTCACAGACATCCCATATCTCAANTAGGAGGTTATGTTATCTTAAAGTC	11128
OY	1501	GTTTAAATGATATATATGCGATACCTGATAGCTTCACATTTTATCTATATCCGTTTA	1560
Db	11127	GTTTAAATGATATATATGCGATACCTGATAGCTTCACATTTTATCTATATCCGTTTA	11068
OY	1561	GATGATTAACATGACATATATCATATTCATTAACGGTTTACAAAACACTGACGATTAATATA	1620
Db	11067	GATGATTAACATGACATATATCATATTCATTAACGGTTTACAAAACACTGACGATTAATATA	11008
OY	1621	TTATCTCTCATATTTGTCACATACGCTTCATTTGTAATCTTATATGGTTCCCTTTCACAATC	1680
Db	11007	TTATCTCTCATATTTGTCACATACGCTTCATTTGTAATCTTATATGGTTCCCTTTCACAATC	10948
OY	1681	TTAACGAAAGTGAATTTACACACATTTTATTTATTCGATACATGATAGTTTGTATATG	1740
Db	10947	TTAACGAAAGTGAATTTACACACATTTTATTTATTCGATACATGATAGTTTGTATATG	10888
OY	1741	AAATCGTGTTTAAACCTTATATGAAACCCCAAGTTTATTCGACCCGATAGCCTTAGSTAAA	1800
Db	10887	AAATCGTGTTTAAACCTTATATGAAACCCCAAGTTTATTCGACCCGATAGCCTTAGSTAAA	10828
OY	1801	TGGGATATGGAACAACAACAGATATATATAGATGTTTGTACAGATCTATAAGAAATATGCA	1860
Db	10827	TGGGATATGGAACAACAACAGATATATATAGATGTTTGTACAGATCTATAAGAAATATGCA	10768
OY	1861	TATGAAGTGAATGGAAGATTTAAATTTGCTTCTGCTGGTATACGAAAAACGCCCTTGAT	1920

Db	10767	TATGAAGTGAAGAGATTAAATGCTTGTGCTATACCGAAAGCGCTTGGAT	10708
QY	1921	ACAAGCGTCATTGTTGAACCTTTGTACGTGACCAATCTTTGAGCGGTGCATTATTGAA	1980
Db	10707	ACAAGCGTCATTGTTGAACCTTTGTACGTGACCAATCTTTGAGCGGTGCATTATTGAA	10648
QY	1981	AACAAATPAAAGTATCTATATATGACAGAGTACATATGCAATATCCGCTPAAACTGAA	2040
Db	10647	AACAAATPAAAGTATCTATATATGACAGAGTACAAATATGCAATATCCGCTPAAACTGAA	10588
QY	2041	ATTGTAATGSGTAAATGATATGATGCAATATTTACTGATGAACTAATATGAACCGTGAA	2100
Db	10587	ATTGTAATGSGTAAATGATATGATGCAATATTTACTGATGAACTAATATGAACCGTGAA	10528
QY	2101	TTTATATTAAGAAGCGCTGAGAAAATTTGCGACATGCTATCTATGATATCTTTAT	2160
Db	10527	TTTATATTAAGAAGCGCTGAGAAAATTTGCGACATGCTATCTATGATATCTTTAT	10468
QY	2161	ATTGAAAGTGACATCGGTCATTCTTCACTTAAAGCATTAATTTCCAGTTGAACGTCAGTA	2220
Db	10467	ATTGAAAGTGACATCGGTCATTCTTCACTTAAAGCATTAATTTCCAGTTGAACGTCAGTA	10408
QY	2221	CATTAACCAATCTATTTGCATATATTAAAGGTGAACATGATGAATPAAAAAAGCGCAAC	2280
Db	10407	CATTAACCAATCTATTTGCATATATTAAAGGTGAACATGATGAATPAAAAAAGCGCAAC	10348
QY	2281	TGTTAA 2286	
Db	10347	TGTTAA 10342	

RESULT 3

ID AAA69032 standard; DNA; 222 BP.

AC AAA69032

DT 27-OCT-2000 (first entry)

DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF017

Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW

XX 3

XX
XX
13030003202E-73

09-TTN-2000
XX
FD

XX 02-DEC-1000. 0000-TP03040
DE

03-DEC-1998. 0815-0110992

PR	03-JUN-1999;	99US-0326144.
PR	28-SEP-1999.	99US-0407804

PR	30-SEP-1999;	99US-0157218.
PR	01-DEC-1999:	99US-0168777

PR	02-DEC-1999;	99US-0454252.
XX		

XX
FA (PHAG-) PHAGETECH INC.
XX

XX
F1
FELICIEL, GROS P, DUBOW M,

DR WEI, 2000-412301/32
DR P-PSDB; AAB16547.

aa	Identifying a bac
PT	

PT inhibits bacteria when a bacteriophage infects a bacterium -

Example 9; Page 277; 456pp; English.

XX

CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides identifying a bacteria-inhibiting function when an
CC uncharacterized bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AA68243 to AA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

XX Sequence 222 BP; 80 A; 30 C; 40 G; 72 T; 0 other;

Query Match 9.7%; Score 222; DB 21; Length 222;

Best Local Similarity 100.0%; Pred. No. 4,4e-27;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1871 ATGGAAGATTAAATGCTCTGCTGATACCGAAACGCGCTTGATACAGCGTCG 1930

DB 1 ATGGAAGATTAAATGCTCTGCTGATACCGAAACGCGCTTGATACAGCGTCG 60

QY 1931 ATTTGAACTTGTACGTAACAATCTTTGACGGTCCCATTTGAAACCAATAAA 1990

DB 61 ATTTGAACTTGTACGTAACAATCTTTGACGGTCCCATTTGAAACCAATAAA 120

QY 1991 GTATCTAATAGACCAAGTACATATCGATATCCGCTAAACGAAATGTATGTC 2050

DB 121 GTATCTAATAGACCAAGTACATATCGATATCCGCTAAACGAAATGTATGTC 180

QY 2051 GTAATGTATATGATGAATATTTTACTGATGACTTAATATGA 2092

DB 181 GTAATGTATATGATGAATATTTTACTGATGACTTAATATGA 222

RESULT 4
AA69048/c
ID AAA69048 standard; DNA; 153 BP.

XX AAA69048;

XX 27-OCT-2000 (first entry)

DE Bacteriophage 44AHD nucleotide sequence 44HDDRF031.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;

KM bacterial growth inhibition; bacterial infection; ds.

OS Bacteriophage 44AHD.

PN WO200032825-A2.

PD 08-JUN-2000.

PF 03-DEC-1999; 99WO-IB02040.

PR 03-DEC-1998; 98US-0110992.

PR 03-JUN-1999; 99US-0326144.

PR 28-SEP-1999; 99US-0407804.

PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.

PR 02-DEC-1999; 99US-0454252.

PA (PHAG-) PHAGETECH INC.

PI Pelletier J, Gros P, Dubow M;

DR WPI; 2000-412361/35.

DR P-PSDB; AAB16563.

PT Identifying a bacteriophage coding region for treating bacterial

PT infections comprises identifying a nucleic acid encoding a product that

PT inhibits bacteria when a bacteriophage infects a bacterium

PS Example 9; Page 278; 456pp; English.

CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides identifying a bacteria-inhibiting function when an
CC uncharacterized bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AA68243 to AA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

XX Sequence 153 BP; 52 A; 21 C; 27 G; 53 T; 0 other;

Query Match 6.7%; Score 153; DB 21; Length 153;

Best Local Similarity 100.0%; Pred. No. 5e-16;

Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1437 TCACATTACTGATGATATCAAGAACCCATCTCAATAGAGAGTATGATCTAA 1496

DB 153 TCACATTACTGATGATATCAAGAACCCATCTCAATAGAGAGTATGATCTAA 94

QY 1497 AGTCGTTTAAATGATTTATATGATACCTGATACGTTTCACTTTAATTATTCG 1556

DB 93 AGTCGTTTAAATGATTTATATGATACCTGATACGTTTCACTTTAATTATTCG 34

QY 1557 TTTAGATGATACATGATGATATCAATATCAT 1589

DB 33 TTTAGATGATACATGATGATATCAATATCAT 1

RESULT 5
AA69065/c
ID AAA69065 standard; DNA; 129 BP.

XX AAA69065;

XX 27-OCT-2000 (first entry)

DE Bacteriophage 44AHD nucleotide sequence 44HDDRF047.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;

KM bacterial growth inhibition; bacterial infection; ds.

OS Bacteriophage 44AHD.

PN WO200032825-A2.

PD 08-JUN-2000.

PF 03-DEC-1999; 99WO-IB02040.

PR 03-DEC-1998; 98US-0110992.

PR 03-JUN-1999; 99US-0326144.

PR 28-SEP-1999; 99US-0407804.

PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.

PR 02-DEC-1999; 99US-0454252.

PA (PHAG-) PHAGETECH INC.

PI Pelletier J, Gros P, Dubow M;

DR WPI; 2000-412361/35.

DR P-PSDB; AAB16580.

PT Identifying a bacteriophage coding region for treating bacterial

PT infections comprises identifying a nucleic acid encoding a product that

PT inhibits bacteria when a bacteriophage infects a bacterium

PS Example 9; Page 279-280; 456pp; English.

CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AA68243 to AA69442 and AAB1523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

XX SQ Sequence 129 BP; 42 A; 17 C; 20 G; 50 T; 0 other;

Query Match 5.6%; Score 129; DB 21; Length 129;

Best Local Similarity 100.0%; Pred. No. 3.5e-12;

Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 589 TCACAACTTAAACAGATTTATATACGATTTTGAATGAATGATGATGAT 648

Db 129 TCACAACTTAAACAGATTTATATACGATTTTGAATGAATGATGAT 70

QY 649 AGTGAACCTATGATGCTGTGAAATGTTTGCAAACTCAGACCTGACCACTTACA 708

Db 69 AGTGAACCTATGATGCTGTGAAATGTTTGCAAACTCAGACCTGACCACTTACA 10

QY 709 TACATTCAT 717

Db 9 TACATTCAT 1

RESULT 6

AAA69066/c

ID AAA69066 standard; DNA; 129 BP.

XX AAA69066;

DT 27-OCT-2000 (first entry)

DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF045.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;

KW bacterial growth inhibition; bacterial infection; ds.

OS Bacteriophage 44AHJD.

PN WO200032825-A2.

PD 08-JUN-2000.

PF 03-DEC-1999; 99WO-IB02040.

PR 03-DEC-1998; 98US-0110992.

PR 03-JUN-1999; 99US-0326144.

PR 28-SEP-1999; 99US-0407804.

PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.

PR 02-DEC-1999; 99US-0454252.

PA (PHAG-) PHAGETECH INC.

PI Pelletier J, Gros P, Dubow M;

DR WPI; 2000-412361/35.

DR P-PSDB; AAB16581.

PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX Example 9; Page 280; 456pp; English.
XX The present invention describes a method for identifying a bacteriophage
XX coding region encoding a product active on an essential bacterial
CC coding region encoding a product active on an essential bacterial

CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AA68243 to AA69442 and AAB1523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

XX SQ Sequence 129 BP; 39 A; 26 C; 21 G; 43 T; 0 other;

Query Match 5.6%; Score 129; DB 21; Length 129;

Best Local Similarity 100.0%; Pred. No. 3.5e-12;

Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1845 TCATAGAAATATGATATGATGAAGTGAATGAATGAATGAATGATGATGAT 1904

Db 129 TCATAGAAATATGATATGATGAAGTGAATGAATGAATGAATGATGATGAT 70

QY 1905 GAAAACGCTTTGATTAAGCGTGTGATTTGAAACCTTTGATGATGATGAT 1964

Db 69 GAAAACGCTTTGATTAAGCGTGTGATTTGAAACCTTTGATGATGATGAT 10

QY 1965 CGGTGCCAT 1973

Db 9 CGGTGCCAT 1

RESULT 7

AAA69070

ID AAA69070 standard; DNA; 123 BP.

XX AAA69070;

DT 27-OCT-2000 (first entry)

DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF051.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;

KW bacterial growth inhibition; bacterial infection; ds.

OS Bacteriophage 44AHJD.

PN WO200032825-A2.

PD 08-JUN-2000.

PF 03-DEC-1999; 99WO-IB02040.

PR 03-DEC-1998; 98US-0110992.

PR 03-JUN-1999; 99US-0326144.

PR 28-SEP-1999; 99US-0407804.

PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.

PR 02-DEC-1999; 99US-0454252.

PA (PHAG-) PHAGETECH INC.

PI Pelletier J, Gros P, Dubow M;

DR WPI; 2000-412361/35.

DR P-PSDB; AAB16585.

PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX Example 9; Page 280; 456pp; English.
XX The present invention describes a method for identifying a bacteriophage
XX coding region encoding a product active on an essential bacterial
XX target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an

CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

SQ Sequence 123 BP; 45 A; 23 C; 11 G; 44 T; 0 other;

Query Match 5.4%; Score 123; DB 21; Length 123;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1562 ATGATACAACTGACATATACATATACATTAACGGTTACAAAACACTGACGTAATATAT 1621
DB 1 ATGATACAACTGACATATACATATACATTAACGGTTACAAAACACTGACGTAATATAT 60
|||||

QY 1622 TATTTCTTACATTTGTACATACAGTTTCATTGTAATTAATTTGGTCTTTCCAACTACT 1681
DB 61 TATTTCTTACATTTGTACATACAGTTTCATTGTAATTAATTTGGTCTTTCCAACTACT 120
|||||

QY 1682 TAA 1684
DB 121 TAA 123
|||

RESULT 8
AAA69077/c
ID AAA69077 standard; DNA: 111 BP.

AC AAA69077;
XX
DT 27-OCT-2000 (first entry)

XX Bacteriophage 44AHJD nucleotide sequence 44HJDORF058.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
KM bacterial growth inhibition; bacterial infection; ds.

XX Bacteriophage 44AHJD.

XX WO200032825-A2.

XX PD 08-JUN-2000.

XX PF 03-DEC-1999; 99WO-IB02040.

XX PR 03-DEC-1998; 98US-0110992.

XX PR 03-JUN-1999; 99US-0326144.

XX PR 28-SEP-1999; 99US-0407804.

XX PR 30-SEP-1999; 99US-0157218.

XX PR 01-DEC-1999; 99US-0168777.

XX PR 02-DEC-1999; 99US-0454252.

XX (PHAG-) PHAGETECH INC.

XX PI Pelletier J, Gros P, Dubow M;

XX DR WPI; 2000-412361/35.

XX P-PSDB; AAB16392.

XX PT Identifying a bacteriophage coding region for treating bacterial
XX infections comprises identifying a nucleic acid encoding a product that
XX inhibits bacteria when a bacteriophage infects a bacterium
XX
XX Example 9; Page 280; 456pp; English.
XX
XX The present invention describes a method for identifying a bacteriophage
XX coding region encoding a product active on an essential bacterial
XX target. The method comprises identifying a nucleic acid sequence encoding
XX a gene product that provides a bacteria-inhibiting function when an
XX uncharacterised bacteriophage infects a pathogenic bacterium. The
XX compound active on a target of a bacteriophage inhibitor protein in a

CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.

SQ Sequence 111 BP; 32 A; 20 C; 18 G; 41 T; 0 other;

Query Match 4.9%; Score 111; DB 21; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1751 TTAACCCCTATTGAACCCCGATTATTCGACCCGATAGCCTTAATGGGATATTG 1810
DB 111 TTAACCCCTATTGAACCCCGATTATTCGACCCGATAGCCTTAATGGGATATTG 52
|||||

QY 1811 AAACGACAGATAGATAGATGTTGTACTGATCATAGAAATATGCAT 1861
DB 51 AAACGACAGATAGATAGATGTTGTACTGATCATAGAAATATGCAT 1
|||||

RESULT 9
AAA69086
ID AAA69086 standard; DNA: 102 BP.

AC AAA69086;
XX
DT 27-OCT-2000 (first entry)

XX Bacteriophage 44AHJD nucleotide sequence 44HJDORF066.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
KM bacterial growth inhibition; bacterial infection; ds.

XX Bacteriophage 44AHJD.

XX WO200032825-A2.

XX PD 08-JUN-2000.

XX PF 03-DEC-1999; 99WO-IB02040.

XX PR 03-DEC-1998; 98US-0110992.

XX PR 03-JUN-1999; 99US-0326144.

XX PR 28-SEP-1999; 99US-0407804.

XX PR 30-SEP-1999; 99US-0157218.

XX PR 01-DEC-1999; 99US-0168777.

XX PR 02-DEC-1999; 99US-0454252.

XX (PHAG-) PHAGETECH INC.

XX PI Pelletier J, Gros P, Dubow M;

XX DR WPI; 2000-412361/35.

XX P-PSDB; AAB16601.

XX PT Identifying a bacteriophage coding region for treating bacterial
XX infections comprises identifying a nucleic acid encoding a product that
XX inhibits bacteria when a bacteriophage infects a bacterium
XX
XX Example 9; Page 281; 456pp; English.
XX
XX The present invention describes a method for identifying a bacteriophage
XX coding region encoding a product active on an essential bacterial
XX target. The method comprises identifying a nucleic acid sequence encoding
XX a gene product that provides a bacteria-inhibiting function when an
XX uncharacterised bacteriophage infects a pathogenic bacterium. The
XX compound active on a target of a bacteriophage inhibitor protein in a
XX bacteria is used to treat or prevent a bacterial infection in an animal.
XX AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
XX nucleotide and protein sequences which are used in the exemplification of
XX the present invention.
XX
XX Sequence 102 BP; 33 A; 16 C; 13 G; 40 T; 0 other;

QY	1300	ATGCATATACGCTGTAATTCGTTGGTATATATGAATGGAATACCTTCATGCAGCTGAT	1359
Db	4599	TATATATATATATATATATATATCTTATTTCTAAACCCCTACATATAAATTTAAAAACTTATA	4540
QY	1360	ATTATTTTTCAAACGATATTTTATTTAAACACAGGTAGCTTAAAAACAAATCAATATG	1419
Db	4539	ATTAATCTT---AAACAAATATAAATTAATAAAACCTTAATTAACATTCATAAAAAATCTA	4483
QY	1420	ACATACACCTTACGACTATACATCTACTGATGATACACAGCAACCCCATACATCAATGAG	1479
Db	4482	AAAAAATATTTATTTTCTATATATATAATTTATTTTCACACCTTCAATATTTAAAAAACA	4423
QY	1480	GAGGTATATGTTATCTAAAGTCGTTTTAAATGATTAATATGACATACCTCATTAACGTTA	1539
Db	4422	TCTCCTCAATATATATTAATAAAACCTTTTATTAATAATAT-TTATATGCAAAATTTAATTC	4364
QY	1540	CATTTTAACTTATTCGCTTTAGATGATACATGACATGACATACATATCATTAACGTTAC	1599
Db	4363	AAAAATTAATTTTCAATATCTATAAATAAATTTAAATTAATAATATATATTAACACAA	4304
QY	1600	AAAAACACGACGAGCTAATATA	1620
Db	4303	ACATATATTTTAAATTAATTA	4283
RESULT 11			
ABL33013/C			
ID	ABL33013	standard; DNA; 7597 BP.	
AC	ABL33013;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE		Human immune system associated gene SEQ ID NO: 986.	
XX			
KM		Human; immune system disease; cytosine methylation; antiasthmatic;	
KM		antiarteriosclerotic; antianaemic; cytosarctic; nootropic;	
KM		neuroprotective; anti-HIV; anticonvulsant; ophthalmological;	
KM		anti rheumatic; antiarthritic; antidiabetic; antipsoriatic;	
KM		antirheumatism; cancer; eye disease; arteriosclerosis; anaemia;	
KM		acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
KM		neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
XX		gene; ds.	
OS		Home sapiens.	
XX			
PN	WO200200928-A2.		
XX			
PD	03-JAN-2002.		
XX			
PE	02-JUL-2001: 2001WO-EP07537.		
XX			
PR	30-JUN-2000: 2000DE-1032529.		
XX			
PPR	01-SEP-2000: 2000DE-1043826.		
XX			
PA	(EPIC-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
DR	WPI: 2002-130909/17.		
XX			
PT		Nucleic acid comprising fragment of chemically modified gene, useful	
XX		for diagnosis and treatment of diseases associated with abnormal	
PT		cytosine methylation	
XX			
PS	Claim 1: SEQ ID NO 986; 32pp + Sequence Listing; German.		
XX			
CC		The present invention provides a number of human immune system associated	
CC		genes which are modified by the methylation of cytosines. The sequences	
CC		can be used in the diagnosis and treatment of immune system disorders,	
CC		including eye diseases such as retinopathy, neovascular glaucoma and	
CC		macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid	

Query Match	4.2%	Score 95.4	DB 24	Length 7597	
Best Local Similarity	45.3%	Pred No. 8.7e-07			
Matches 472	Conservative	0	Mismatches 561	Indels 10	Gaps
CC	leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,				
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel				
CC	diseases. The present sequence is a gene of the invention.				
XX					
SQ	Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 other;				
QY	251 CAAAACAGATATTATCATGATGTCACATACCTGTAATATAATAGCATATCATTTTTCAC	310			
DB	1322 CAAATATATCATCTCAACCTCTAAACCTAAAAATATATATATATATATATATATAT	1263			
OY	311 TTTAAGCACACCTGCGTTATTTTGATATATATTTACACGCAAAATATATATTTAAATCTG	370			
DB	1262 ATAAAT	1203			
OY	371 CAGAGAGAAATGGAACACACATCTAAAAATGAGAGGCTCTATTTTAGCCAAAAATCTAA	430			
DB	1202 ATAAAT	1143			
OY	431 ATGTATATTTTGGAAAAACGCTTAAATCTTCAATCAATTTACATTTTAAATGTTTTAA	490			
DB	1142 ATAAAT	1083			
OY	491 ATGGTTTAAATTTTAAATTTATGTGATACCTTTTGGAAACCATCATCATGTCACAT	550			
DB	1082 ATACTAT	1023			
OY	551 TAGTAGAATTAATCTCT--GATGGTGTTAATTAACAGATACACACTTAAACAGATT	607			
DB	1022 TACAT	963			
OY	608 TTAATTTATACATTTTGTGATAAAGATATATATATATATATATATATATATATATATAT	667			
DB	962 TAAAT	903			
OY	668 CTGGAATATGTTTGGCAAACTCAGACCTGACACACTTACATATCTATATATGACGTGA	727			
DB	902 TAAAT	843			
OY	728 TTAATTTATGATATGTCATATTCATATATATATATATATATATATATATATATATAT	787			
DB	842 TTAATACAT	783			
OY	788 AATATACATTTTTCATGAAATATATATATATATATATATATATATATATATATATAT	847			
DB	782 AAT	723			
OY	848 AGTTATCTACCA----ATATACAGATATATATATATCTTATACATTTATCATTTCCA	902			
DB	722 AAT	663			
OY	903 TGAATGAAATTTTATGACAT	962			
DB	662 AAT	603			
OY	963 CACCAATATCAT	1022			
DB	602 AAT	543			
OY	1023 TTAATGAT	1082			
DB	542 AAT	485			
OY	1083 TTTACAGAACCAAGTTATCCCTACTTTTATAGATGATGACATTTATTTTCTATATAT	1142			
DB	484 ATAAAT	425			
OY	1143 GATTAT	1202			
DB	424 ATAAAT	365			

Thu Jan 9 14:50:10 2003

us-09-727-892a-2.rn1

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 01:18:41 ; Search time 70 Seconds
(without alignments)
10015.187 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 2286
Sequence: 1 atgggattctagatgcattc.....taaaaaaggaactgttaa 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 862724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCRUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	85	3.7	19124	2	US-08-487-826B-13
C 2	78	3.4	6768	1	US-08-107-755A-1
C 3	78	3.4	8457	1	US-07-991-867B-1
C 4	78	3.4	8457	2	US-08-544-332-1
C 5	78	3.4	8457	4	US-09-370-861A-1
C 6	70.6	3.1	19124	2	US-08-487-826B-13
C 7	67.8	3.0	827	4	US-08-998-416-535
C 8	66.4	2.9	1431	4	US-09-316-083-2
C 9	66.2	2.9	665	4	US-08-998-416-937
C 10	66	2.9	1511	1	US-07-991-867B-8
C 11	66	2.9	1511	1	US-08-107-755A-8
C 12	66	2.9	1511	2	US-08-544-332-8
C 13	66	2.9	1511	4	US-09-370-861A-8
C 14	65.6	2.9	837	4	US-08-998-416-188
C 15	64.6	2.8	615	4	US-08-998-416-186
C 16	63.8	2.8	834	4	US-08-998-416-305
C 17	63.8	2.8	3701	4	US-08-845-258-10
C 18	63.8	2.8	3701	4	US-08-990-571-10
C 19	63.8	2.8	3701	4	US-08-723-142A-10
C 20	63.8	2.8	3701	4	US-09-528-784A-10
C 21	63.6	2.8	662	4	US-08-998-416-185
C 22	63.6	2.8	767	4	US-08-998-416-172
C 23	63.4	2.8	854	4	US-08-998-416-534
C 24	63.2	2.8	724	4	US-08-998-416-683
C 25	63	2.8	732	4	US-08-998-416-1036
C 26	62.8	2.7	990	4	US-09-134-001C-283
C 27	62.2	2.7	828	4	US-08-998-416-538

C 28	61.6	2.7	636	4	US-08-998-416-1137	Sequence 1137, App
C 29	61.4	2.7	636	4	US-08-998-416-1137	Sequence 1137, App
C 30	61.4	2.7	711	4	US-08-998-416-786	Sequence 786, App
C 31	61.4	2.7	837	4	US-08-998-416-288	Sequence 288, App
C 32	60.8	2.7	665	2	US-08-883-795A-36	Sequence 36, App
C 33	60.8	2.7	701	4	US-08-998-416-701	Sequence 701, App
C 34	60.8	2.7	2960	3	US-08-913-842-3	Sequence 3, App
C 35	60.4	2.6	1511	1	US-07-991-867B-8	Sequence 8, App
C 36	60.4	2.6	1511	1	US-08-107-755A-8	Sequence 8, App
C 37	60.4	2.6	1511	2	US-08-544-332-8	Sequence 8, App
C 38	60.4	2.6	1511	4	US-09-370-861A-8	Sequence 8, App
C 39	59.8	2.6	663	4	US-08-998-416-191	Sequence 191, App
C 40	59.4	2.6	168575	4	US-09-426-290-1	Sequence 1, App
C 41	59	2.6	20674	4	US-09-641-638-651	Sequence 651, App
C 42	58.8	2.6	168575	4	US-09-426-290-1	Sequence 1, App
C 43	58.4	2.6	6124	4	US-08-213-419B-3	Sequence 3, App
C 44	57.4	2.5	4810	3	US-08-852-629-11	Sequence 11, App
C 45	57.2	2.5	860	4	US-08-998-416-287	Sequence 287, App

ALIGNMENTS

RESULT 1
US-08-487-826B-13/C
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: SLM, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhan
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
City: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13
Query Match 3.7%; Score 85; DB 2; Length 19124;
Best Local Similarity 43.2%; Pred. No. 1.3e-07;


```

; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: Saliwanchik, David R.
;   REGISTRATION NUMBER: 31,794
;   REFERENCE/DOCKET NUMBER: 0F114.C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
;   INFORMATION FOR SEQ. ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 6768 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: unknown
;     MOLECULE TYPE: DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM: Amsacta moorei entomopoxvirus
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: complement (65..1459)
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 1474..2151
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: complement (2239..2475)
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 2502..2987
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 3080..6091
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: complement (6277..6768)
;     US-08-107-755A-1
;
Query Match      3.4%; Score 78; DB 1; Length 6768;
Best Local Similarity 42.1%; Pred. No. 2.2e-06;
Matches 573; Conservative 0; Mismatches 785; Indels 4; Gaps 2;
```

```

QY 258 AGATATTATCATGATGACATACATGTAATTAATACGATAATCTTTTACTTAAGA 317
   |||||
DB 1599 AATATTTATTTATTTATGTTAAATTTGTCATAAATATATATCTGATTAAT 1640
QY 318 CACCATGCGTATTTTGTATATATTACACCGAATAATATATTAAATCTGCAGAGA 377
   |||||
DB 1639 TTAATTTATTTATTTGTTTATTTAGTTAGTATATGATTAATATATATTATTAT 1580
QY 378 AATGACACACATTAATAATGAAGAAGGCTACTATTTTACCCAAAATCAAAATGTA 437
   |||||
DB 1579 TATTTATTAATAAATCTATATTAACACAAATGTTAATATATAATACCTATTAAC 1520
QY 438 TTTCAGAAAAGCGTATTAATCTTCATCAATTTAGATTTAACATGTTTAAATGTTT 497
   |||||
DB 1519 ATATTAATAATTTATTAATAAATGTTGAATTAAGTAAAGCAATTTATTAATAATA 1460
QY 498 TAAATTTATATTTTGTGATACCTTTATGAAGAACCAATACATCAATTCACACATT 557
   |||||
DB 1459 ATGAA-TAATTAATTAATGAAATTTCCAAATTAATAAATTAATGACAGATCTG 1401
QY 558 GAATTTACTGATGCTGTTTATTAACAGATTCACACTTAATAACAGATTTTATTTAC 617
   |||||
DB 1400 TAAATTTATCTCTATGTTATTTTATAGTAAATAGATAATATGTTATTTTAAATCC 1341
QY 618 GATTTTGTGATAAGATATG--ATATGAATAGTAGAAGCCCTATGACTATGCTGTA 674
   |||||
DB 1340 TATTAATAATATATCTAATGAAATATAGCTATTTAGAAAAAATAGATGATGATGA 1281
QY 675 ATGTTTTCGAACACTCAACCTGAAACAACTACATACATTCATATATGACGATTATAT 734
   |||||
DB 1280 AACAAAGAGTAAGTAACTTATATACCTATTAATAATCTTGGAATTTATTAATTA 1221
```

```

QY 735 AGGATATGCGCATATTCATTAATAGATATATTTCCAAATTTTGACTATACAAATTAAC 794
   |||||
DB 1220 AGTTTATGATCCAAATCATATTTATATATTTATTTTGAATAATATTTAAAGAAAATT 1161
QY 795 ATTTCATTAATTAATTAATGGAATCTTACTGGAATAATGACACAGCTTTCACTTACT 854
   |||||
DB 1160 TTTTCATTTTGGACTATATAAAAAATATTACAGATTAATAATATTAATTAATCTTT 1101
QY 855 CAACCATATCAAGATATTAATAATATCTTATACACTTATCATTTCCATGATATGTAATT 914
   |||||
DB 1100 TATTAACATTAATAAATAATTAAGCAATTCATTAATATATGAGAAATCTTTCTTA 1041
QY 915 TTATGACTATATTAATCATATCTATGCTGTTGTTAAATATGATATACCAATATACAT 974
   |||||
DB 1040 ATCAGATTAATCCAAACAGAGATTTTCTACATTTGCCACATTTATGATGATTGGA 981
QY 975 AAACAACTATATGATGAGCCTTTGTTTCTATTCATCAATCAATTCGATTAATCTTATGT 1034
   |||||
DB 980 TGCTCAATAGAAATACCCGCTTTGGATTTATCCTAAACCTAAAGCTTACCAGAAA 921
QY 1035 GATGTATCAATGAATAATTCACACATGTTATATCTTTTACGAACTATTCAGAACCAAC 1094
   |||||
DB 920 ACATGATTAATCTACATGTTATGTTGCGGACATATATTAATTTGATATATCCATTAC 861
QY 1095 GTTAATCCCTACTTTTATGATGATGACATTAATTTTTCATTAATATAGATTGATPAAGA 1154
   |||||
DB 860 TATGATTAATACAGATTTATATTTTTCATTTGATTTATTTGAAAAACGATGATAA 801
QY 1155 TGTATTTACGATGATTTATTAATTAATTAATACCGTATTTACGTAATGATGTGT 1214
   |||||
DB 800 TATTGCTACAGTACGTTCACTATGAGATGCTTAACCTGGAAGATAGATGAATTTT 741
QY 1215 AATATCTATTAATTAATGATATGATATGATTAATATCAATACAAATCAATTAAGATAT 1274
   |||||
DB 740 TTTAATGAATGATTAATAAAGATTTTATGTTTCTATTAATTAATTAAGATCATTTTAC 681
QY 1275 TCAAGACATTAACGGTATTTGATTCATGCAATATACGTGTTATTCGTTGTTATATATGA 1334
   |||||
DB 680 TTGTTGCGTAAATGATAAACAACCTTGTATGAATAAAGACGCACTTTTATATG 621
QY 1335 ATGTGATATCTTCATGACAGGTATATTTATTTTCAAACTATTTTAAACACAAAG 1394
   |||||
DB 620 TATGCTTATATCCAGACTTTATAAACAATAAATAATATATGTTTATGCAATCTGA 561
QY 1395 TAAATTAATAAACAATAATCAATATGACATCACCTTACGACTACATTAAGATGAT 1454
   |||||
DB 560 TATGACTATTAATAAGCAATTAACACTATAGTACTCTCTAATTAATATATGCTTATTT 501
QY 1455 CACGGAACCCATACGCAATGAGAGGTTATGTTATCAATGCTGTTTAAATGAT 1514
   |||||
DB 500 ATACATTTGATGCTCTACGAAATTTTAATGATATTTAAAGTAATTAATTTACT 441
QY 1515 ATATGCAATACCTGCAATTAAGCTTCAATTTTAACTTATTCGTTTGAATGATTAACATGA 1574
   |||||
DB 440 TTTTAACTTTTGAATTAACATATGATAGTCCGATTTGAGTGTATTAATTAAT 381
QY 1575 ACTATACATATCATTAACGTTTCAAAAAACACTGAACGTA 1616
   |||||
DB 380 TTTATATTAATATGTTTATTTCAACATTAATCAATTAATTTGA 339
```

RESULT 3
US-07-991-867B-1/C
; Sequence 1, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1
City: Gainesville
STATE: FL USA
COUNTRY: FL USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
AUTORNEY/AGENT INFORMATION:
NAME: Salikanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI14.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8457 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: *Amsacta moorei* entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (65..1459)
FEATURE:
NAME/KEY: CDS
LOCATION: 1474..2151
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2239..2475)
FEATURE:
NAME/KEY: CDS
LOCATION: 2502..2987
FEATURE:
NAME/KEY: CDS
LOCATION: 3080..6091
FEATURE:
NAME/KEY: CDS
LOCATION: complement (6277..6768)
US-07-991-867B-1
Query Match 3.4%; Score 78; DB 1; Length 8457;
Best Local Similarity 42.1%; Pred. No. 2.3e-06;
Matches 573; Conservative 0; Mismatches 785; Indels 4; Gaps 2;

Db 1579 TATTTATTAATAAATCTATTATAACAAATGTAAATATTAATACCTATTAACCA 1520
Qy 438 TTTAGAAAAGCGTGTAAATCTTCATCATTTAGATTTAAACAATGTTTTAAATGCTTT 497
Db 1519 ATAAATAAATTTATTTAAAAATGATTTGAATTAACCTAGAAACATTTATTTAAATA 1460
Qy 498 TAAATTAATATTATTGATTAACCTATGAAAACCAATCAATCAATTCGACATTAAGTAA 557
Db 1459 ATGAA-TAATAAATTAAGATTTCCAAATATAAATTTAAAAATGCCAGATCTGGTAT 1401
Qy 558 GAAATTAAGTGGTGTATTATTACAGAAATCAACACTTAATAAAGATTTAATATATAC 617
Db 1400 TAAATTTATGCTATGTTATTTTATAGTAAATATGATATATGTTATTTAATAATCC 1341
Qy 618 GATTTTATTAAGTAATG---ATATGATGATAGTAAACCTATGCTATGCTGGTGA 674
Db 1340 TATTAAATATATTAATGATTAACCAATATAGCTATATTAGAAAAATATAGATGATGA 1281
Qy 675 ATGTTTGCACAACTCACACCTGGAACAATTACATCAATTCATTAAGACGTATTAAT 734
Db 1280 AACAGAGCTAAAGTAACATTTTATACCTATTAATTAATCTTGGAATATTAATGAT 1221
Qy 735 AGGTATGCGCATATTCATTAATGATATATTTCCAAATTTTGACATTAACAATTAAC 794
Db 1220 AGTTTATGATCCAAATCAATTAATTAATTAATTTGAATTAATTAAGAAAAATTT 1161
Qy 795 ATTTTCATTAATATTATGAATCTTACTGTAATTAAGAAATGACAGCTTTTCAGTACT 854
Db 1160 TTTCTTATTTGGACTATTAATAAATATTACAGATTAATAATTAATATTAATCTT 1101
Qy 855 CAACCAATATCAAGATTTAAATATCTTATACACATTAATTCATTCATATGAATTT 914
Db 1100 TATTACATCTAATAAATAATTAAGGATTCATTTGATATATGAAAGCTTTTCTAAA 1041
Qy 915 TTATGACTATATTAATATCTATTCGCGTGTTAAATGATTAACCAACAATTCAT 974
Db 1040 ATCAGATTAATCCAAAGAAAGATTTTTCACATTTGCCACATTAATGATGATTTGA 981
Qy 975 AAACAATTAATGATGAGCTTGTGTTTCTATTGACATCAATTCGATTAATCTTATGT 1034
Db 980 TGCTCAATATAGAAATACCGCTTTGGATTTAACTTAACCTAAAGCTTAACCCAGAA 921
Qy 1035 GATGATCATGAATAAATTCACATGCTTATCTTTAGCAACATTCAGAACCAAC 1094
Db 920 ACATGATTAATCTACATGCTTAAGTCGGAGATATATATATGATTAATTCATTAAC 861
Qy 1095 GTTAATCCCTACTTTTATGATGATGACATTAATTTTCATTAATTAAGATTAAGA 1154
Db 860 TATGATTAATACAGATTAATGATTTTTCATTTGATTTTGAATAAAGCTGATAAAA 801
Qy 1155 TGTATTTAGCATGATTAATTAATTAATAATTAATCAAGCTGATTAAGTCAAAAGATGT 1214
Db 800 TATTGCTACAGTACCTTCATTAAGAGTCTAATAACTTGAATAGAGTAATTTT 741
Qy 1215 AAAATCTATTAATATGATTAATGATTAAGTAAATATCAATCAATCAATTAAGAT 1274
Db 740 TTTATGAATGATTAATAAAGATTTTATGTTTCCATTAATTAATGATCACTTTTAC 681
Qy 1275 TCAAGACATTAAGGATTTGTTGATGATGATTAAGCTGATTAATGCTTTGATTAATGA 1334
Db 680 TTTGTCCTATTAATTAACACTTTGATTAAGATTAATAAAGCTGATCACTTTTATTAAT 621
Qy 1335 ATGTAATTAATTTATGACAGCTGATTAATTTTCAAACTATTTTAACACAGAG 1394
Db 620 TAGTGGTATTAATACAGACTTAATAAACAATAAATAATTAATGTTTGAATGTA 561
Qy 1395 TAACTTAATAAACAATCAATTAATGATCACTTACGACTATCAATTAATGATAT 1454
Db 560 TATGACTTAATAAAGCATAAACAATCAATTAATGATCACTTAATTAATGATTTAT 501
Qy 1455 CAAGAACACCCATTAACCAATTAAGAGGATTAATTAATTAATGATTTTAATGATTT 1514
Db 500 ATACATGATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 441

QY 1515 ATATGGCATACGTCATTACGTTACATTTTAACTTATCCGTTTAGATGATAAATGA 1574
Db 440 TTTTAATCTTTGATATGATAGTCCGATTTGGTATATATATATAT 381
QY 1575 ACTATACATATCATTAACGCTTACAAAACACGTAACGTAA 1616
Db 380 TTTATATATATAGTTTATTTTCAACATATAAATCTAATTTGA 339

RESULT 4
US-08-544-332-1/c
Sequence 1, Application US/08544332
Patent No. 5935777

GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gerard H. Bengen
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/544,332
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,867
FILING DATE: 07-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,755
FILING DATE: 19-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA: US 07/657,584
APPLICATION NUMBER: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bengen, Gerard H.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: UFI14.C4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8457 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (65..1459)
NAME/KEY: CDS
LOCATION: 1474..2151

FEATURE:
NAME/KEY: CDS
LOCATION: complement (2239..2475)
FEATURE:
NAME/KEY: CDS
LOCATION: 2502..2987
FEATURE:
NAME/KEY: CDS
LOCATION: 3080..6091
FEATURE:
NAME/KEY: CDS
LOCATION: complement (6277..6768)
US-08-544-332-1

Query Match 3.4%; Score 78; DB 2; Length 8457;
Best Local Similarity 42.1%; Pred. No. 2.3e-06;
Matches 573; Conservative 0; Mismatches 785; Indels 4; Gaps 2;

QY 258 AGATATTATCATGATTCGACATACGTTATTAATACGATATCATTTTACTTAAAGA 317
Db 1699 AAGGATTTATTTATTTATGTTAAATTTGTCAAAATATATATCTGATATATGATTAAT 1640
QY 318 CACCATGCGTTATTTGATATATATTACAGCGCAAAATATATTTAAATCTGCAGAGA 377
Db 1639 TTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1580
QY 378 AATGGAACACATTTAAATGAAAGGCTACTATTTTACCCAAAATCAAAATGTAAT 437
Db 1579 TATTTATTAATAAATCTATATTTAAACAAATGTTATATATAAATCCATATATCCAA 1520
QY 438 TTGAGAAACGTTAAATCTCAATCAATTTAGATTTAACAATGTTTAAATGGTT 497
Db 1519 ATATATAAATTTATTTAAATATGATTTGAATTAACHTGAACATTTTATTAATAATA 1460
QY 498 TAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 557
Db 1459 ATGAA-TAAATAAATTTAGAAATTTCCAAATTAATAAATTTAAATTTCCGAATCTGTAT 1401
QY 558 GAATTTACTTGATGGTGTATTTAAAGATGACAACTTAAACGATTTTATATATAC 617
Db 1400 TAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1341
QY 618 GATTTTGTATTAAGATTAAG--ATATGATGATGATGAGGCTATGACTATGCTGTGA 674
Db 1340 TATTAATATATTAATGACAAATATAGCTATATATAGAAAATATGATGATGATGA 1281
QY 675 ATGTTTGCAAAACCTCACACCTGACAACTTACATACATTCATTAAGCGTATTAAT 734
Db 1280 AACAGAGGTTAAAGTAACTTTATACCTATTAATATCTTGAAATTTATATATATGAAT 1221
QY 735 AGGTATGCGCATATTCATATATGATATATTTCCAAATTTGACTATTAACAAATTAAC 794
Db 1220 AGTTTATGATCCAAATCATATTAATATATTTTGAATAATATTAAGAAAT 1161
QY 795 ATTTTATGATATTTATGATCTTACTGAAATATGAAAGACACGTTTACGTTACT 854
Db 1160 TTTTCTATTTTGGACATTAATAAATATTTTACAGTAAATAATATTAATATTAATCTTT 1101
QY 855 CACCAATATCAAGATATTAATAATATCTTATACATCTTATTCATGATATGATAT 914
Db 1100 TATTTATCTMAAATATTAATAAGCATTTCCATAGATATATAGAGAAATCTTTCTA 1041
QY 915 TTATGACTATTTTAAATCAATTTCTATGCTGGTGTAAATATGATATACCAAAATACAT 974
Db 1040 ATCAGATTTATCCAAAACAGAGATTTTCTATCATTTGCCCAATTTATGATATTTGA 981
QY 975 AACCAAGCAATTTGATGAGCGCTGTTTCTCTATGSCATCAATTTGAGTATTCCTAT 1034
Db 980 TGCTCAATATGGAATACCGCTTTTGGATTTATCTTAACCTTAAGCTTACCCAGAA 921
QY 1035 GATGTATCATGAAAAATTTCCAAACATGTTATATCTTTTACGAACATTTACGAACCAAC 1094
Db 920 ACATGATAAATCTACATGTTAAGTCGGAGATATATATATTTGATATATTCATTAAC 861

QY 1095 G T A A P C C C T A C T T T T T T T T A C A T A T G C A C A T A T T T T T C A T A T A T A T A T A T A T G A T A A G A 1154
 Db 860 T A T A T A T A A T A A C A G A T A T A G A T A T T T T C A T 801
 QY 1155 T G T A T T T A A C G A T A T T T A T T A T T A A A T T A A A T C A C G T A T A T A C G T C A A T G A T T G T 1214
 Db 800 T A T T G C T A C A G T A G C T C A C T A C T A N G A T A G C T A T A A C T T G A A G A T A G A G T A A A A T T T T T 741
 QY 1215 A A A A T A C T A T A T A T A T A T A T A T A T A C T T A T A T C A T A C A A T A C T A T T A G A A T G A T 1274
 Db 740 T T T A T A A A T A G A T A A A A A A A A G A T T T T T A T G T T C C T A A A T T T T A A T A G A T C A T T T T A C 681
 QY 1275 T C A A G A C A T T A C G G T A T T T A T T A T G C A T A T A C G A T A T A C G T A A T T T G T T G T A T A T A T A G A 1334
 Db 680 T T G T T G G T A A T A A G A T A A A C A C T T G A T A A A G A T A A A A A A G C T G C A T A C T T T T T A A T A A G 621
 QY 1335 A T G T G A A T A C T T T C A T G A C G A C G T A T A T T A T T T T C A A A C T A T T T A T T A T A A A C A A G 1394
 Db 620 T A G T G G T A T A T A T A C C A A A C T T A T A A A C A A A T A A A A A T A T A G T T A T T A T T A T T A C A T C G A 561
 QY 1395 T A A T T A A A A A C A A A A T C A A T A T A T A T A G A C A T A C C T T A C G A C T A C A T A C T A G A T A T 1455
 Db 560 T A T A C T A T A A A A A A A A C C A T A A A C A C T A C A T A T A C T C C T A T A C T A A T A T A T A T T A T T 501
 QY 1455 C A A G A A C C C A C T A C T A C A A T A G A G A G G T T A T G T A T C T A A A G T C G T T T A A A T G A T T 1514
 Db 500 A T A C A T T A G A T A C T A T A C A G A A T A T T A A T A A T A T T A A A A A A G T A A A T T A T T A C T 441
 QY 1515 A T A A G C A T A C T G C A T T A C T A C T T C A C A T T T A C T A T T T C G T T A G A T A G A T A A C A A A G A 1574
 Db 440 T T T T A A T A C T T T G A A T T A C A A A T A G A T A G T A G T C C G A T T G T G A T A C T T T A A T A T A T A T T 381
 QY 1575 A C T A T A C A A T A C A T T A A C G G T T A C A A A A A C A C T A G A C G T A A 1616
 Db 380 T T T A T A T A T A T A G T T A T T A T T C A A C A A T A A A T C T A A A T T T G A 339

 RESULT 5
 US-09-370-861A-1/c
 : Sequence 1, Application US/09370861A
 : Patent No. 6410221
 : GENERAL INFORMATION:
 : APPLICANT: Moyer, Richard W.
 : APPLICANT: Grulid, Michael E.
 : TITLE OF INVENTION: No. 6410221e1 Entomopoxvirus Expression System
 : FILE REFERENCE: UFI14.C4.D1
 : CURRENT APPLICATION NUMBER: US/09/370,861A
 : CURRENT FILING DATE: 1999-08-09
 : PRIOR APPLICATION NUMBER: US 07/991,867
 : PRIOR FILING DATE: 1992-12-07
 : PRIOR APPLICATION NUMBER: US 08/107,755
 : PRIOR FILING DATE: 1993-08-19
 : PRIOR APPLICATION NUMBER: WO 92/14818
 : PRIOR FILING DATE: 1992-02-12
 : PRIOR APPLICATION NUMBER: US 07/827,685
 : PRIOR FILING DATE: 1992-01-30
 : PRIOR APPLICATION NUMBER: US 07/657,584
 : PRIOR FILING DATE: 1991-02-19
 : NUMBER OF SEQ ID NOS: 78
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 1
 : LENGTH: 8457
 : TYPE: DNA
 : ORGANISM: Amsacta moorei entomopoxvirus
 : US-09-370-861A-1

258 AGATATTATCATGATTGCACATACTGTATAATAATACGATAATCATTTTCTACTTAAGA 317

Db	1699	AAAGATTATATTATTTAGTTAAATTTGGTCAAAAAATATATATACGTGAATATCGTATAAT	1640
Oy	318	CACCATCGGTAATTTTGGATATATATTCACGGCAAAATATATATTTAAATCTGCAGAGA	377
Db	1699	TTAAATTTATTTATATGTTATATTTATTTGTCGCTTATATGATATATATATATTTTATAT	1580
Oy	378	AAATGACACACATTTAAAAATGAAAGGCTACTATTTAGCCAAAAATCAAAATGTAAAT	437
Db	1579	TATTTATTTAAAAATCTATTTTAAACAAAGTATATATATATATATACCTTATATACCA	1520
Oy	438	TTTAGAAAAAGGTGTAATCTTCAATCAATTTAGATTTAACATGTTTTTAAATGCTTT	497
Db	1519	ATAATAAAAATTTATTTAAAAATATGTTGTAATAAATGAAACATTTTATTTAAATATA	1460
Oy	498	TAAATTTATATTTATGATTAAGTACTTTATGAAAACCAATCATCAATTTGCAATTAGGTA	557
Db	1459	ATGAA-TAATTAATTTGAAAGATTTCCAAATTTAAAAATTTAAAAATGCCAGAACTGGTAT	1401
Oy	558	GAATTAAGTATGATGGTGTATTTTAAACAGAAATCACACTTAAACAGATTTTAAATATAC	617
Db	1400	TAAATTTATGATATGTTATTTTATTTTGTAGTAATATAGTAAATATGTAATTTTATTAATCC	1341
Oy	618	GATTTTGGATTAAGATATAG--ATATGATATAGTAGAGGCTATGCTATGCTGGA	674
Db	1340	TATTAATATAAATACATATCCAAATTAACCTATATTTAGTAAAAAATAGATGATGATGA	1281
Oy	675	ATGTTTGCAAAATCTCACCTGACAGAACTTACATCATCATCTTAATGACGATATATTT	734
Db	1280	AACAGAGGTAAAGTATACATTTATACCTTATATAATACCTGGAAATATATATATATGATTT	1221
Oy	735	AGTATGTGCCATTTTCATTTATATGATATATTTCCAAATTTTGACTATTAACAATTTAAC	794
Db	1220	AGTTTGTAGATCAAAATCATATTTATATATTTTATTTTGGAAATATATTTAAAGAAAAAT	1161
Oy	795	ATTTTCATGTAATTTATGGAATCTTACTGATGATATGAAATGACAGCTTTTCAGTACT	854
Db	1160	TTTTCTATTTTGGACTATAAAAAAATATTTACAGATAAAAATATATAATATATACTTT	1101
Oy	855	CAACCAATATCAGATATTTAAAAATATCTTATACACTTATCTATTTCCATGATGAAATTT	914
Db	1100	TATTAATCTTAAAAAATATATATAGGCAATTCATTAATATATGAAAGTCTTTTCTTAA	1041
Oy	915	TTATGACTATATTAATCATTTATGTGTGTGTTTAAATATGATATTAACCAATATCAT	974
Db	1040	ATCAGAAATTTCCAAAAACAAGAGATTTTCTCATTTGSCCAATTTTATGATGATTTGGA	981
Oy	975	AAACAACATATATGATAGCGCTGTTTTCATATTCAGATCAATTCAGATATACCTATATG	1034
Db	980	TGCGCAATATGAGATATACCGCTTTGGGATTTATCTTAACCTAAAGCTTACCCAGAA	921
Oy	1035	GATGATATCAGAAAAAATTTCCAAATGATGTTATATCTTTTCGAGACCTATTCAGAACCAAC	1094
Db	920	ACATGATATAAATCTACATGTTTAAAGTTCGGAGATATATATATTTATATATATTCATTTAA	861
Oy	1095	GTTATATCCATCTTTTATATGATGACAAATATTTTATATATATATATATATATATATATAT	1154
Db	860	TATGATTTAATACAGATTTATGATTTATTTTCAATTTTATTTTATTTGAAAAACGTATAAAA	801
Oy	1155	TGATTTTACAGATGATTTATTTAAATTTAAATCACGTTATTTAGCTCAATGATTTG	1214
Db	800	TATTTGCTACAGTACTCTCATCTATGATGATGATATTAACCTTGAAGATTAAGATTAATTTT	741
Oy	1215	AAATACTATATATATGATATATGATTTACGTTATATATCATATCAATATATATATATATAT	1274
Db	740	TTTATATGAATGATAAAAAAGATTTTTTATATGTTTCTATATATTTAAATATGATCATTTTAC	681
Oy	1275	TCAGACATTTACGGGTTTATGATGCATGATATATAGCTTTATTTGTTATGTTATATATAT	1334
Db	680	TTGTTGGTATATATATATAACCTTGATATATAAGATAAAAAGCTGATACCTTTTATATAG	621
Oy	1335	ATGGAATACTTTCATGACAGGTATATATTTTTCAAAACTATTTTATTTAAAAACAAGG	1354

Thu Jan 9 14:50:10 2003

us-09-727-892a-2.rni

Page 7

Db	620	TAGGGTTATATACGCACTTTATAACCAATATAAATATATGTTTATTGAACTCGA	561
QY	1395	TAAGTAAAAAACAATAATCATATGACATCACCTTAGCATATCATTAAGATAT	1454
Db	560	TATGACTATATAAAGCCATTAACACATCAATAGTACTCTTAATTAATATATGCTTATT	501
QY	1455	CAAGCAACCCCATCTCAATCAGGAGGTATGTTATCTAATAGTCGTTTAAAGATT	1514
Db	500	ATACATTGATGTACTATACAGAAATATTTAATGATATATTTAAAAATGMAATTATCTT	441
QY	1515	AATAGCATACCTGCATCTAGTGCATTTTAACATTATCCGTTAGATGATCAATCA	1574
Db	440	TTTTAATACGTTTGAATTAACAATATGATAGTCCGATTTGGTATGTTTAATATAATT	381
QY	1575	ACTATACAAATCATTTACGGTTTACAAAAACACTGAGCTAA	1616
Db	380	TTTATATATATAGTTTATTTCAACATTAATAATTTAATTGCA	339

```

RESULT 6
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF INVENTION: 45
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEITICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13
;
Query Match 3.1%; Score 70.6; DB 2; Length 19124;
Best Local Similarity 43.4%; Pred. NO. 5.8e-05;
Matches 569; Conservative 0; Mismatches 734; Indels 8; Caps
120 AACCTTACTATTCCTGACCAATGGTGGTTAATGCTTATGCAATTAATGATGTTGAGT 179

```

Dd	5792	AAGGGTATACGTTTTTTTTTTTTTTTTTTTGTCATATATATAAAATAGTAACCTACA	5851
Oy	180	ATTTCGGATTTCGAATCCTTTTGTAGCGATTTCAGTAGTGGAAAACGCTGATAC	239
Dd	5852	ATATGCATATTACAGAAGATATATTTGGTAAAAAATATATATATATATATATAAGA	5911
Oy	240	AATCACAAAATCAAAAACAATA--TTATCAGATTGCCACTACTGTAATTAATACGAT	297
Dd	5912	CATTAAGCATATACATACTAGTAATAGTTATAGTTTATATATACATCCCTTTATATATTAAT	5971
Oy	298	AATCATTTTACTTAAAGACACCATGCGCTATTTTGTATATATATACCGGAAAATTA	357
Dd	5972	TTTTTTGTTTACTCTTGTCGCTCTTTTGTATATATATATACCAATATATAAACAA	6031
Oy	358	TATTTAAATCTGCAGAGAAGAAATGACACACATTTAAAATGAAGAGCGTACTATTTTA	417
Dd	6032	TAT	6091
Oy	418	GCCAAAATCAAAATGTATTTAGAAAACGTTTAATCTTCATCAATATTTGATTTA	477
Dd	6092	TAT	6150
Oy	478	ACAATGGTTTAAAGSTTTTAAATTTAAATATATATGATTAACCTTATGAACCAATACA	537
Dd	6151	TGCATTTTATATATTTTACTATATACCTTAAAGATATATTAATATTTATATAGTACGA	6210
Oy	538	TCAATTGCACATATTAGTAAAGAAATTACTTGATGGTGTTATTTAACGAATCACACTT	597
Dd	6211	TATATGATTTAT	6270
Oy	598	AAAACAGTTTATATTTACGATTTTGTAAAGATATGATATGAATGAAGATAGGAAGCC	657
Dd	6271	TTTATATATACATCATTTTGAATATATATATTTATATAGTATTTTACTATTTTTT	6330
Oy	658	TATGACTATAGCTGTGAATGTTTGGCAAACTGCACGCTGAGCAACCTTACATTCAT	717
Dd	6331	TATATTTATACATTAATTTTGAATTCATTAATCAATGCAATGATTCATTAATAATACAA	6390
Oy	718	AATGAC-CIGATTTATTAGAGTAGGAGCCCATCTCTTTACGATATATTTCCAAATTT	776
Dd	6391	CAAACCAATATGATBAACACTTTTATATATATATATATATATATATATATATATAT	6450
Oy	777	TGACTATTAACAAATTTAATCATTTTCATGTAATTTATGSAATCTTACTGGAATATGAAT	836
Dd	6451	TTCCGTTATTTATTTATTCATTTTTTTTTTGTGATGCTATATATATATATATATAT	6510
Oy	837	GACACGTTTCAGTTACTCAACCAATATCAAGTATTAATAATCTTATACACATTATCA	896
Dd	6511	ATATATATATACAACAAAATTTATATATATATATATATATATATATATATATATAT	6570
Oy	897	TTTCCATGATATGATTTTATTTAGCT--AATATTAATCTTATGCTGGGTTTAAT	954
Dd	6571	TACAAAGATATGATCTATATCATATATATATATATATATATATATATATATATAT	6630
Oy	955	ATGATATACCCAATATCATATAAACCAATATATGATGAGCCCTGTTTTCATTTGACATC	1014
Dd	6631	TATAGATAGAGAGAAAGAAAGACATATATGCTCTTTGTTATCTTAATATATATATA	6690
Oy	1015	AATTCAGATATCTTTTGATGATGTCATGAAAAATTCACACATGGTTATACTTTAC	1074
Dd	6691	TATATATATATTTAAATTAAGTCAAAAAAATATCATATTTATGTATATATATTA	6750
Oy	1075	GACACATTTGAGAACCCAGCTTAATCCCTACTTTTTAGATGATGACAAATATTTTICA	1134
Dd	6751	TATATTAACCGTGCATATATACCTTTTATATGTTTATGTTATTCGATTTTTTTTTTC	6810
Oy	1135	TTATATAGATGATTAAGTATGATTTTAACGATTTTATATTAATTAATTAACACGT	1194
Dd	6811	TCATTTATATATTTACTTATATATATATATATATATATATATATATATATATATAT	6870
Oy	1195	GTTATACGTCAAATGATTGTAATATCTATATATATGATTAAGATATAGTTATATCAT	1254
Dd	6871	AGATATATTAAGGAATTCATTAATATATATATATATATATATATATATATATATAT	6930

QY	168	TCATGTCGAGTATTCGCGATTCGCAATCTTTTATGACGCATTTTATACGATGTGAA	227
Db	165	TAAATATAAATTTTATTATTATTATGATGATGATTAATGAATATTAATCAACATCATAA	224
QY	228	AAGAGCGTATGCAATCACAAATGCAAAACGATATATCATGATTCGCATACGTGA	287
Db	225	ACAAGCTACGCTATGCTATTAAGAGTTATGCAATTCGAAAAATTTTGTGATTATA	284
QY	288	TAAATACGATATCATCTTTTACTTAAAGACACCAGCTGTTATTTGATATATTCACG	347
Db	285	ATTATTATATCTGATTTTATTTTAAATGAATGAATGAATGAATGATATATATA	344
QY	348	CGAAATATATATTTAAATGTCGAGAGAAATGACACACATTTAAATGSAAGAGCG	407
Db	345	TAAATATATATATATATTTCTATTTAAATATATATGAATTTTAAAAATATATGATATA	404
QY	408	TACTATTTTACGCAAAAAATCAAAATGTAATTTAGAAAAAGCTTTAAATCTTCAATCA	467
Db	405	TTTAAATTTTAAACCTGCTGTAATTTGAAACATAATTTCTAAATATTTTATTATGA	464
QY	468	TTTACATTTAAACAATGTTTTAAATGTTTTAAATTAATATATGATACCTTATGA	527
Db	465	TAAATATTTAA-----TTAATAAATATATGAAATATTTAGATATGTTAAATATGATTCT	519
QY	528	AACCAATGATCAATTCGACATTAGGTAGAAATAATCTGAGTGCTGTTATTTACAGA	587
Db	520	AATATTTATATGTTTAAATATATTAATTAATTAAGSTAAATTAAT-----ATTAAACA	573
QY	588	ATCCAACTTTAAACAGATTTTATTTATACGATTTTGATAAAGATATGATATGATGA	647
Db	574	GATTTAGATTTAAATTAATTAAGAATTTATGATTTATTAACAGGTTAATTGAG--GTGA	632
QY	648	TAGTAGAACCTTATGATCTATGCTGGAATGTTTTGGCAAACTCACACGCTGAACCTTAC	707
Db	633	TGTTATATTTGCTGCTGAGGTATTCGAATTTACTATACATGCTATGATGATTTAATAC	692
QY	708	ATACATCATATGACGATGATTAATTTAGTATGTGCCATATTCATTAATGATATATTT	767
Db	693	TATCTTTTATTAATAAGAAATTAATAATAGATTTTGTAGAAATGATGATCTCTT	752
QY	768	TCCAAATTTTACATATACAAATTAACAATTTTCATGATATATATGAAATCTTACTGA	827
Db	753	AAAAGATATCTCTTATTTTGTAAATGCTTTCTCTATTAATATTAATAACTATTTAGCTAA	812
QY	828	TATGAAATGACAGCTTTTCAGTTACTGCAACCAATATCAAGATTTTAAATATCTTATAC	887
Db	813	AGAAAAGATTTTCTATATATTTTATTAATTAATTAATGATTAATTAATTAATCAAT	872
QY	888	ACATATCATCTTCATGATGATGATTTTATGCTATATTAATTCATTCCTATGCGGTG	947
Db	873	TAAATATCATATCCCTTATTTATTAATTTTAAAAATTAATTAATAATTTACCTTATAAA	932
QY	948	TTTAAATATGTAATACCAAAATACATAAACAACATTAATGATGAGCCTGTTTTCTAT	1007
Db	933	TATATGGAATTTTAAATATATTTATGTTAGCTGTTTACGCTGACGATGTTCTTT	992
QY	1008	TGACATCATATCGATTAACCTTATGATGATGATGATGAAAAATTCACAATGTTATA	1067
Db	993	TTTATCATCTATGTAATATCTTAAGATACATTAATTTAAAAAATATGACCTAGTTA	1055
QY	1068	CTTTAGCAACATTTTCAGACCAACGTTAACTCCCTACTTTTTTATGATGATGACATATA	1127
Db	1053	TGTTATTTTCAACAGTGAACACGTTAAGATTAATTTTATTTATTCAGSANTCTTTTCA	1113
QY	1128	TTTTTCAATATATAGATGATAAGATG-----TATTTAACGATGATTAATTAAT	1176
Db	1113	TTTATCATTTTCTATATTTTAAAAAAGTGTGTAATGAAAAATTAAGATTTTAAATTAAT	1172
QY	1179	TAAATTTAAATCAGCTATATAGCTCAAAAGTGTGTAATTAATCTATATATGATGAATGA	1236
Db	1173	TACAGACATGATGATTAATTAATGAATTTATTT-----TATATTTTGTATTAATTTTATAC	1222
QY	1239	TTACGTTAATATCAATCAAAATACATTAGCATATGATTCAGACATTCAGGATATTTGATG	1296

[illegible]

```

1      RESULT 9
2      US-08-998-416-937
3      Sequence 937, Application US/08998416
4      Patent No. 6239264
5      GENERAL INFORMATION:
6      APPLICANT: Philippsen, Peter
7      APPLICANT: Pohlmann, Rainer
8      APPLICANT: Steiner, Sabine
9      APPLICANT: Mohr, Christine
10     APPLICANT: Wendland, Jurgen
11     APPLICANT: Knechtle, Philipp
12     APPLICANT: Rebschuhung, Corinne
13     TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
14     TITLE OF INVENTION: AND USES THEREOF
15     NUMBER OF SEQUENCES: 1152
16     CORRESPONDENCE ADDRESS:
17     ADDRESSEE: NO.6239264atlis Corporation
18     STREET: 3054 Cornwallis Road
19     CITY: Research Triangle Park
20     STATE: NO. 6239264th Carolina
21     COUNTRY: USA
22     ZIP: 27709
23     COMPUTER READABLE FORM:
24     MEDIUM TYPE: Floppy disk
25     COMPUTER: IBM PC compatible
26     OPERATING SYSTEM: PC-DOS/MS-DOS
27     SOFTWARE: Patentln Release #1.0, Version #1.30
28     CURRENT APPLICATION DATA:
29     APPLICATION NUMBER: US/08/998,416
30     FILING DATE: 24-DEC-1997
31     CLASSIFICATION: 435
32     PRIOR APPLICATION DATA:
33     APPLICATION NUMBER: CH 0016/97
34     FILING DATE: 31-DEC-1996
35     ATTORNEY/AGENT INFORMATION:
36     NAME: Melgs, J. Timothy
37     REGISTRATION NUMBER: 38,241
38     REFERENCE/DOCKET NUMBER: P/5-30306/A/CGC1976
39     TELECOMMUNICATION INFORMATION:
40     TELEPHONE: 919-541-8587
41     TELEFAX: 919-541-8689
42     INFORMATION FOR SEQ ID NO: 937:
43     SEQUENCE CHARACTERISTICS:
44     LENGTH: 665 base pairs
45     TYPE: nucleic acid
46     STRANDEDNESS: single
47     TOPOLOGY: linear
48     MOLECULE TYPE: DNA (genomic)
49     ORIGINAL SOURCE:
50     ORGANISM: PAG1581RP
51     IS-08-998-416-937

```

Query Match	2.98;	Score 66.2;	DB 4;	Length 665;
Best Local Similarity	47.48;	Pred. No. 0.00027;		
Matches 313;	Conservative	0;	Mismatches 328;	Indels 20;
			Gaps	3;
QY	304	TTTTACTTAAGACACCATCGCTTATTTTGATATATATTACACGGCAAAATATATATTTA	363	
Db	3	TCTTATTTTAAATTTTATTTATTAACATTTTATTAATTTAGAAAATATATATCTAGAGATATA	62	
XY	364	AAATCTCAGAGAAATGAAACACNCATTAATAATGGAAGAGGGCTACTATTTTACCCAA	423	

Page 11

FILING DATE: 30-JAN-1992
Prior APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI14.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus

FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-08-107-755A-8

Query Match 2.9%; Score 66; DB 1; Length 1511;
Best Local Similarity 43.6%; Pred. No. 0.00032;
Matches 499; Conservative 0; Mismatches 630; Indels 15; Gaps 4

Dy 260 ATATTTATCATGTCACACATACTGTAAATTAAGTATATCATTTTTTACTTAAGACA 319
| | | | |
Db 379 ATTATTTCTACTAGGAATTAATTTGATATAGAATTAATTAATTTCTGTAAAGCA 438
| | | | |
Qy 320 CCATCGCTATTTTGGATTAATTTACACGCAGAAAATATATTTAAATCTGCAGAGANA 379
| | | | |
Db 439 CAATTTATCCAGCACATTAACCTTTTTTTATATTAGCATTTTATCCAAAAATGTG 498
| | | | |
Qy 380 ATGAACACACATTTAAAAATGAAAGSGGTACTATTTTGCCAAAAATCAAATGTAAATT 439
| | | | |
Db 499 TCTAATATATTTTCTCTCAAAAATTTGACACTCTATATGCCATATATATCATAAATATCT 558
| | | | |
Qy 440 TAGAAAAACGCTGTTMAATCTTCATCAATTTAGATTTAACATGTTTTAAATGGTTTA 499
| | | | |
Db 559 ACGATATGATTTTCATTAATTAATTTATTTGTTTAAATGATTAATATCTTATTTAAT 618
| | | | |
Qy 500 AATTAAATATTATGATTAACCTTTATGAAGAACCAATACATTCATTCGACATTAGTAGA 559
| | | | |
Db 619 ATAATTCGCTAGATTTATTAATATTTTATTTATTAATATTAATTT--ATCATATTTAGA 675
| | | | |
Qy 560 AATTACTTGATGGGTGTTATTTTAACAGATGACAACTTMAAACAGATTTTATTTATAGA 619
| | | | |
Db 676 GTATTAATTTACACATTTTGGATTAGATTAATAATATATTCATTAATTTTTCGACATCAATCT 735
| | | | |
Qy 620 TTTTGTGATTAAGATTAATGATATGATGATAGTAGAAGCCCTATGACATATGCTGTGAATGTT 679
| | | | |
Db 736 GTTGTTTGGCCAGAAAACATAGGACCAATTTAATTTTATATGACATTTTTTTTTTATATAT 795
| | | | |
Qy 680 TTGC AAAACCTCACACCTGMAACACTTACATTCATTCATATAGAGCGTGATATATTAGTGA 739
| | | | |
Db 796 TTGATATATTTTTCAAAAAAAATTAATTCATATGAAAAAAAATTAATTTATCCAATAGG 855
| | | | |
Qy 740 TGTCGCAATTCATTAATGATATATTTCCAAATTTTGACTATTAACAATTAACATTTT 799
| | | | |
Db 856 ATTTACTTAATTTGATTAATTAATTTAAATTAATTTTAAATTAATTAATTTAAAA---A 912
| | | | |
Qy 800 CATGAATATTAATGAAATCTTACTTGAAATTAATGAATGACACGTTTTCAGTTACTACACC 859

QY 740 TGTGCCATATTCATTATAGATATATTTCCAAATTTTGACTATACAAATTAACATTTT 799
Db 856 ATTACTATAATTCGATATAATTTTAAATATTTTAAATATTTTAAATTTTAA--A 912
QY 800 CATGAATATATGGAATCTTACTGATATAGAAATGACACGTTTCAGTTACTACACC 859
Db 913 AAATTAATTAATACAGAGATATATGTTATTAATTTTAAATTAATTAATTAATTTATTT 972
QY 860 AATATCAAGATATTAATATCTTTATACATTTATCCATTTCCATGATATGATTTTATG 919
Db 973 TAGAAGATTCATATATATATATATGATATATATTTTAAAT-----AATATCCAG 1026
QY 920 ACTATATTAATCATTTCTATGCTGTGTTTAAATATGATATACACCAATATCAATAACA 979
Db 1027 AAAATATTAAGATTTATATTTTCCAAATTTTAAATATATTTTAAATTTTAAACAA 1086
QY 980 AACTAATATGAGCCTTGTCTTCTAT--TGACATCAATTCAGTTATCCCTTATGTA 1036
Db 1087 AATTAATAAATATACATATTTAGATATATCTTTATACAAAATAGCAATATATAGTATA 1146
QY 1037 TGTATCATGAAAAATTCACACATGTTTATCTTTACGAACTATTCAGAACCAACGT 1096
Db 1147 TTATACACCATCTTATAGAAATTTTAAATTTGATCATGTAATATATTAATGACTATA 1206
QY 1097 TAATCCCTACTTTTATGATGATGACAAATTTTTCATATATATAGATGATTAAGATG 1156
Db 1207 ATTTTATTAATTAATTTTATTAATTTTAAATTAATTAATTAATTAATTTTGTGA 1286
QY 1157 TATTTAAGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGTA 1216
Db 1267 ACTTATTAATGTTTTCCTATATGTTATGTTAGTTAGTTAAATGGAATCAATTAACAAATTA 1326
QY 1217 AATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1276
Db 1327 AAGTTTAAATTTATAGAAAAATTAATTAATTTTAAATTAATTAATTAATTAATTTTCAATG 1386
QY 1277 AAGCAATTAAGCTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1336
Db 1387 TTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1446
QY 1337 GTGATATCTTATGACGATGATTTATTTTCAAACTATTTTATTAATTAATTAATTAATTA 1396
Db 1447 ATCATATATTAAGAAATTAATTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATG 1506
QY 1397 AGTT 1400
Db 1507 AATT 1510

RESULT 13
US-09-370-861A-8
: Sequence 8, Application US/09370861A
: Patent No. 6410221
: GENERAL INFORMATION:
: APPLICANT: Moyer, Richard W.
: APPLICANT: Hall, Richard L.
: APPLICANT: Grubel, Michael E.
: TITLE OF INVENTION: No. 6410221el Entomopoxvirus Expression System
: FILE REFERENCE: UFI14.C4.D1
: CURRENT APPLICATION NUMBER: US/09/370,861A
: PRIOR APPLICATION NUMBER: US 07/991,867
: PRIOR FILING DATE: 1992-12-07
: PRIOR APPLICATION NUMBER: US 08/107,755
: PRIOR FILING DATE: 1993-08-19
: PRIOR APPLICATION NUMBER: WO 92/14818
: PRIOR FILING DATE: 1992-02-12
: PRIOR APPLICATION NUMBER: US 07/827,685
: PRIOR FILING DATE: 1992-01-30
: PRIOR APPLICATION NUMBER: US 07/657,584
: PRIOR FILING DATE: 1991-02-19
: NUMBER OF SEQ ID NOS: 78

SOFTWARE: PatentIn version 3.1
: SEQ ID NO 8
: LENGTH: 1511
: TYPE: DNA
: ORGANISM: Amsacta moorei entomopoxvirus
US-09-370-861A-8
Query Match 2.9%; Score 66; DB 4; Length 1511;
Best Local Similarity 43.6%; Pred. No. 0.00032;
Matches 499; Conservative 0; Mismatches 630; Indels 15; Gaps 4;
QY 260 AATATATCATGATGACATATGATATTAATTAATTAATTAATTAATTAATTAATTAATTA 319
Db 379 ATTTTCTTCTACTTATGAAATTAATTTGATATAGAAATTAATTAATTTTCTGTTAAAGTGA 438
QY 320 CCAATGCTTATTTGATATATTTACACGCGAAATATATTAATTTAAATCTGCAGAAAGAA 379
Db 439 CAATTTAATCCACACAAATTAATTTTATTTTATTTATGACATTTTATCCAAATTTGT 498
QY 380 ATGACACACATTTAAAAATGAAAGGCTACTATTTTACGCAAAATCBAATGTATTT 439
Db 499 TCTAATCATTTTCTTCAAAAATTTGACACTCTATGCAATTAATTAATTAATTAATCT 558
QY 440 TAGAAAACGTTTAAATCTTCAATCAATTTAGATTTTAAACAAATGTTTAAATGCTTTA 499
Db 559 ACGATATGATTTTCAATTAATTAATTTTATTTTATTTATGATATTAATTTCTTATTAAT 618
QY 500 AATTAATATTAATTTGATTAATTTATGAAACCAATCAATCAATTTGCAATTTAGTAA 559
Db 619 ATATTTCCGATGATTTATTTATTTTATTTTATTTTAAATCTATTT--ACTATATATATGA 675
QY 560 AATTAATGAGGTTTATTTAAGAAATCAACACTTAAACAGATTTAATTAATTAATGA 619
Db 676 GTTATTAATTAACATTTTATTTATGATTAATTAATTAATTAATTTTGTGCAATTTCT 735
QY 620 TTTTGTAAAGATTAATGATATGATATGATAGTAAAGCTTATGCTGTAATTTCT 679
Db 736 GTTGTTTGCGAAGAAACATGAGCAATTAATTAATTTATTTTCAATTTTATTTTATTA 795
QY 680 TTGCAAAACACACCTGCAACCTTATGATTAATTAATTAATTAATTAATTAATTAATTA 739
Db 796 TTTATATTTTTCACAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 855
QY 740 TGTGCCATATTCATTATAGATATATTTTCCAAATTTTGACTATACAAATTAACATTTT 799
Db 856 ATTACTATAATTCGATATAATTTTAAATATTTTAAATATTTTAAATTTTAAATTTTAA--A 912
QY 800 CATTTGAATTTATGGAATCTTACTGATATAGAAATGACACGTTTTCAGTTACTACACC 859
Db 913 AAATTAATTAATTAACAGAGATATGTTATTAATTTTAAATTAATTAATTAATTTATTT 972
QY 860 AATATCAAGATATTAATTAATCTTTATACATTTATCCATTTCCATGATATGATTTTATG 919
Db 973 TAGAAGATTCATATATATATATATGATATATATTTTAAAT-----AATATCCAG 1026
QY 920 ACTATATTAATCATTTCTATGCTGTGTTTAAATATGATATACACCAATTAACATAACA 979
Db 1027 AAAATATTAAGATTTATATTTTCCAAATTTTAAATATATTTTAAATTTTAAACAA 1086
QY 980 AACTAATATGAGCCTTGTCTTCTAT--TGACATCAATTCAGTTATCCCTTATGTA 1036
Db 1087 AATTAATAAATTAATACATTTATGATATATCTTTATACAAAATAGCAATATATAGTATA 1146
QY 1037 TGTATCATGAAAAATTCACACATGTTTATCTTTACGAACTATTCAGAACCAACGT 1096
Db 1147 TTATACACCATCTTATAGAAATTTTAAATTTGATCATGTAATATATTAATGACTATA 1206
QY 1097 TAATCCCTACTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1156
Db 1207 ATTTTATTAATTAATTTTATTAATTTTAAATTAATTAATTAATTAATTTTGTGA 1286
QY 1157 TATTTAAGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGTA 1216

This Page Blank (uspto)

Query	Target	Score	DB	Length	35913
1	FILE REFERENCE: 38-10(15810)B	43.98	0.0077	13	2
2	CURRENT APPLICATION NUMBER: US/09/754,853A	0	500	13	2
3	CURRENT FILING DATE: 2001-01-05	0	500	13	2
4	PRIOR APPLICATION NUMBER: US 60/174,880	0	500	13	2
5	PRIOR FILING DATE: 2000-01-07	0	500	13	2
6	NUMBER OF SEQ ID NOS: 1119	0	500	13	2
7	SEQ ID NO 2	0	500	13	2
8	LENGTH: 335913	0	500	13	2
9	TYPE: DNA	0	500	13	2
10	ORGANISM: Glycine max	0	500	13	2
11	FEATURE:	0	500	13	2
12	NAME/KEY: CDS	0	500	13	2
13	LOCATION: (45163)..(45114)..(45450)..(45509)..(46941)..(48763)..(48975)..(49573)	0	500	13	2
14	OTHER INFORMATION: Clone ID: 240017_region_g3	0	500	13	2
15	US-09-754-853A-2	0	500	13	2
16	Query Match	3.0%	Score 68:	DB 9:	Length 335913
17	Best Local Similarity	43.98	Pred. No. 0.0077		
18	Matches 401	Conservative	0	Mismatches 500	Indels 13
19					Gaps 2
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					
87					
88					
89					

Db 99665 AAAAAAAAAAGCTTTAAAGCTATGAGAAAAAAACTTACACTTAAAGTGTAGTT 99724
 QY 1538 CACATTTTAACTATTCGCTTTAGATGATAACATGACACTATACATATGATTAACGGTT 1597
 Db 99725 CAAAGAAAAAAGCTGTACGACTTTAAAAATGTAATATAAATAAATAAGCTAAAGTT 99784
 QY 1598 ACAAAACACGTGAA 1611
 Db 99785 GTAAAAAAATATAA 99798

RESULT 3
 US-09-754-853A-3
 ; Sequence 3, Application US/09754853A
 ; Publication No. US20030005491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hauge, Brian M.
 ; APPLICANT: Parnell, Laurence D.
 ; APPLICANT: Parsons, Jeremy D.
 ; APPLICANT: Wang, Ming Li
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 ; FILE REFERENCE: 38-10(15810)B
 ; CURRENT APPLICATION NUMBER: US/09/754,853A
 ; PRIOR FILING DATE: 2001-01-05
 ; PRIORITY APPLICATION NUMBER: US 60/174,880
 ; PRIOR FILING DATE: 2000-01-07
 ; NUMBER OF SEQ ID NOS: 1119
 ; SEQ ID NO 3
 ; LENGTH: 335913
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (46798)..(48763)..(48975)..(49573)
 ; OTHER INFORMATION: Clone ID: 240017_region_G3
 ; US-09-754-853A-3

Query Match 3.0%: Score 68; DB 9; Length 335913;
 Best Local Similarity 43.9%: Pred. No. 0.0077;
 Matches 401; Conservative 0; Mismatches 500; Indels 13; Gaps 2;

QY 704 TTACATCATTCATGACGTGATTTATTTAGTGTGCGCATTCATTATTAAGTGATA 763
 Db 98892 TTTTATCTTTTATTTCAATCTTTTAAAAATATAAATATTTATTTATTAATTA 98951
 QY 764 TATTTCCAAATTTGACATATACAAATTTTCAATTTGATGATATTAATGATGATGAT 823
 Db 98952 TTAATATAATTTTATTTTACTAGTATTAATGATTTCTTAAATAT-----ATTTCGT 99004
 QY 824 TGAATATGAATGACAGCTTTGATGATGATGATGATGATGATGATGATGATGATGAT 883
 Db 99005 ATATTTAAATTTTATATTTTCTAATGATGATGATGATGATGATGATGATGATGATGAT 99064
 QY 884 ATACCATTTATTCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 943
 Db 99065 AAATATAAATTTTACATATGATATTTATTTATTTTCTTTTAAATTTTAAATTTTAA 99124
 QY 944 GTGGTTTAAATGATATACACCAATTTACATTAACCAATTTGATGATGATGATGATGATGAT 1003
 Db 99125 AAAAAGCTAAATTTTAAATAAAGTTAAATAATTTTGTAGCTAATATTAATTTAAATTTT 99184
 QY 1004 CTATTTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063
 Db 99185 ATGAAATATATTTATTTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 99244
 QY 1064 TATCTTTTACGA-----ACACTATTCAGAACCAAGCTTATCCCTCTTTTATGATG 1117
 Db 99245 AATATATAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 99304
 QY 1118 ATGACATTTATTTTCTATTTATTAATGAATGATGATGATGATGATGATGATGATGATGATGAT 1177

Db 99305 ATAAAAATTTCTTAGTGTCTATTTATTTGTCAGGACGCTTATACATTTATGTTAAT 99364
 QY 1178 TTAATTTAATGACGTGATTTACGTCAAATGATTTGTAATATCTATATATGATATG 1237
 Db 99365 ATACAAATTTTATTTAATTTATTTAATTTATCTCATTTTAAATTTACAAATATTTTAAAT 99424
 QY 1238 ATTACGTTATTCATTCATTAATTCATTAATTCATTAATTCATTAATTCATTAATTCATTAAT 1297
 Db 99425 AAACATTTATTTAATAAATATTTAATAATTTTAAATTTAATAAATTTAATAAATTTAATAA 99484
 QY 1298 GCATGATATATGCTGTTAATTCGTTTGTATATGATGATGATGATGATGATGATGATGATGAT 1357
 Db 99485 AAATATCTATTAATAAATTTACATCAAAATTTTAAATTTAACTAATTAAGTAAATTTAA 99544
 QY 1358 ATATTTATTTTCAAACTATTTTATTAACACAGATGATGATGATGATGATGATGATGATGATGAT 1417
 Db 99545 AATATTTATTTAATTAATTAATTAATTAATTTTATTTCAACTTTTAAATTTTAAATTTGAAA 99604
 QY 1418 TGACATCACCTTACGACTATACATTTACATGATGATGATGATGATGATGATGATGATGATGAT 1477
 Db 99605 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 99664
 QY 1478 AGGAGTTATGTTATCTAAGTCGTTTAAATGATGATGATGATGATGATGATGATGATGATGAT 1537
 Db 99665 AAAAAAATGACCTTTAAAGTCATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 99724
 QY 1538 CACATTTTAACTATTCGCTTTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1597
 Db 99725 CAAAGAAAAAAGCTGTACGACTTTAAAAATGTAATATAAATAAATAAATAAATAAATAAATAA 99784
 QY 1598 ACAAAACACGTGAA 1611
 Db 99785 GTAAAAAAATATAA 99798

RESULT 4
 US-09-754-853A-4
 ; Sequence 4, Application US/09754853A
 ; Publication No. US20030005491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hauge, Brian M.
 ; APPLICANT: Parnell, Laurence D.
 ; APPLICANT: Parsons, Jeremy D.
 ; APPLICANT: Wang, Ming Li
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 ; FILE REFERENCE: 38-10(15810)B
 ; CURRENT APPLICATION NUMBER: US/09/754,853A
 ; PRIOR FILING DATE: 2001-01-05
 ; PRIORITY APPLICATION NUMBER: US 60/174,880
 ; PRIOR FILING DATE: 2000-01-07
 ; NUMBER OF SEQ ID NOS: 1119
 ; SEQ ID NO 4
 ; LENGTH: 513509
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (111805)..(113968)..(114684)..(115204)
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(513509)
 ; OTHER INFORMATION: unsure at all n locations
 ; OTHER INFORMATION: Clone ID: 318013_region_A3
 ; US-09-754-853A-4

Query Match 2.9%: Score 67; DB 9; Length 513509;
 Best Local Similarity 44.7%: Pred. No. 0.012;
 Matches 386; Conservative 0; Mismatches 470; Indels 7; Gaps 3;

QY 424 AATCAAAATGATATTTTGAAGAAAAAGTGTTAAATCTTCAATTCATTTAGATTTAACAATG 463
 Db 464571 AAAAAATTTCTCACATCATTAATGTTTATATATTTACTTTTAAATTTATATATATATATA 464630

```

QY 484 TTTTAAATGGTTTAAATTTAAATATATATGATTAACCTTTATGAAACCAATACATCAATT 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464631 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 464690
QY 544 GCAACATTAGGTAGAAATACCTGATGCTGCTATTTTAAACAGAAATCACACTTAAACA 603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464691 TACATTTATTAACCAATAGATATATATATATATATATATATATATATATATATATAC 464750
QY 604 GATTTTATATAGATTTTGTATTAAGATATATATATATATATATATATATATATATATAT 663
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464751 AATTTTAAATGT--ATATCTATTTATATATATATATATATATATATATATATATATAT 464805
QY 664 TATCTGTGAATGTTTTCGAAACACACCTGACACACTTACATACATTCCTCAATGAC 723
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464806 TACGATATATAGCTATTTTAAATATATATATATATATATATATATATATATATATATTA 464865
QY 724 GTGATTTATAGTATGTCACATTCATTTAGTATATATATATATATATATATATATATAT 783
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464866 TATATTTTATAGTATATATATATATATATATATATATATATATATATATATATATAT 464925
QY 784 AACAAATTAACATTTTCATTTGAATATATATATATATATATATATATATATATATATAT 843
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464926 ATACATGATTCATATATATATATATATATATATATATATATATATATATATATATAT 464985
QY 844 TTTCAGTTACTCAACCAATATCAAGATATATATATATATATATATATATATATATATATAT 903
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464986 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 465045
QY 904 GATATGATTTTATGATATATATATATATATATATATATATATATATATATATATATAT 961
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465046 TATGTAAGAAATATCTCTTCAATATATATATATATATATATATATATATATATATATAT 465105
QY 962 ACACCAATATCAATTAACCAATATATATATATATATATATATATATATATATATATAT 1021
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465106 ATAGCTATATTTTATATATATATATATATATATATATATATATATATATATATATAT 465165
QY 1022 GTTATCTTATGATGATATATATATATATATATATATATATATATATATATATATATAT 1081
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465166 TATATTCATATATATATATATATATATATATATATATATATATATATATATATATAT 465225
QY 1082 ATTCAGACCAACGTTATATATATATATATATATATATATATATATATATATATATATAT 1141
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465226 AAGACTATACCTTATAGATATATATATATATATATATATATATATATATATATATATAT 465285
QY 1142 AGATTTGATTAAGATATATATATATATATATATATATATATATATATATATATATATAT 1201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465286 CATATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 465345
QY 1202 GTCATATGATTTAAATATATATATATATATATATATATATATATATATATATATATAT 1261
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465346 TAAATGTTATATATATATATATATATATATATATATATATATATATATATATATAT 465405
QY 1262 CATTAGAAATGATTCAGACAT 1284
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465406 ATATAGAAATATCTATTTATT 465428

```

```

RESULT 5
US-09-774-414-2
; Sequence 2, Application US/09774414
; Patent No. US20020102231A1
; GENERAL INFORMATION:
; APPLICANT: The Institute of Physical and Chemical Research
; TITLE OF INVENTION: Endonuclease
; FILE REFERENCE: PH-651
; CURRENT APPLICATION NUMBER: US/09/774,414
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 09/306,970
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

```

```

; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1428)
US-09-774-414-2

Query Match
Query Similarity 2.9%; Score 66.4; DB 10; Length 1431;
Best Local Similarity 43.9%; Pred. No. 0.0082;
Matches 543; Conservative 0; Mismatches 671; Indels 24; Gaps 5;

QY 168 TGATGTGAGTATTTCCGAGTTGGATCTTTTATATAGACGATTTATATAGATATGAA 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 224
QY 228 AAGCGTGTATACACCAATATCAAAATCAATATTTATATATATATATATATATATATATAT 287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 ACACGTACACCTATATGCTATATATATATATATATATATATATATATATATATATATATAT 284
QY 288 TAAATACATATATATATATATATATATATATATATATATATATATATATATATATATAT 347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 344
QY 348 CGAAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 404
QY 408 TACTATTTTGGCAAAATATCAAAATATATATATATATATATATATATATATATATATATAT 467
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 TTTAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 464
QY 468 TTTAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 TAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 519
QY 528 AACCATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 587
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 573
QY 588 ATCAACTTAAACAGATTTTATATATATATATATATATATATATATATATATATATATATAT 647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 GTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 632
QY 648 TAGTAAACCTATGACCTATGCTATGTAATGTTTGCACCAATCTACACCTGACCAACTTAC 707
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 TGGTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 692
QY 708 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 767
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 TATCTTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 752
QY 768 TCCAAATTTTGCATTAACCAATTTATATATATATATATATATATATATATATATATATATAT 827
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 AAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 812
QY 828 TATGAAATGACACGTTTTCAGTTACTCAACCAATATATATATATATATATATATATATATAT 887
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 AGAAATGATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 872
QY 888 ACATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 947
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 932
QY 948 TTTAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1007
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 992
QY 1008 TGACATCATTTGAGATATCTTATATATATATATATATATATATATATATATATATATATAT 1067
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 TTTATCATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1052
QY 1068 CTTTATAGACATATATATATATATATATATATATATATATATATATATATATATATATAT 1127

```

```

Db 1053 TGGTATTTCCAGAGTGAAGACAGTGAAGAAATTAATTATTATTATTCAGAAATCTTTTGA 1112
QY 1128 TTTTTCATTTATAGATGATGAAGATG-----TATTACAGATGATTATTAT 1178
Db 1113 TTTATCTATTCTTAAGTTTAAAAAGTTGGTAATAGAAATTTAAAGATTTTAAATTATT 1172
QY 1179 TAAATTTAAATCAAGTGTATTACGCAAAATGATGTAATTAATATATGATATGA 1238
Db 1173 TACAGAACTAGTGAATGAATTAATGAATTTATT---TATTATTGATTAATTTTAC 1229
QY 1239 TTACGTAATATCAATATCAATATCAATATCAATATCAATATCAATATCAATATCAAT 1298
Db 1230 TTTACATGATTAATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATC 1289
QY 1299 CATGATATACGTTGTTATTTGTTATATGATGATGATGATGATGATGATGATGATGATGATG 1358
Db 1290 ATATTAATTTGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1349
QY 1359 TATTATTTTCCAAACTATTATTATTAAACACAGCTA 1396
Db 1350 AATTGATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1387

```

RESULT 6

```

US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

```

Query Match 2.9%; Score 65.4; DB 10; Length 640681;
Best Local Similarity 49.9%; Pred. No. 0.022;
Matches 253; Conservative 0; Mismatches 241; Indels 13; Gaps 3;

```

```

QY 761 ATATATTTCCAAATTTGACATATTAACAATTAACATTTTCATGAAATATGGAATCTT 820
Db 325011 ATATATTTGATTAATTTGATTAATTAATTAATTAATTTTGTTTTAAATTTTAT 325070
QY 821 ACTGGAATTAAGAAATGACAGCTTTGAGTACTCAACCAATATCAAGATATTAATAAT 880
Db 325071 TTTTGTAAATTTGATGAAGCTATTTCTTTAAATTA--AGTAAATTTTAAATTTAA 325128
QY 881 CTTATACACTATTCATTTCCATATATGATTTTGTGACTATATTAATCATCTATC 940
Db 325129 ATTAACGTATTTTATTCATATTATAT---TTTTTAAATTTTAAATTTTCTTAA 325185
QY 941 GTGCTGTTTAAATATGATTAACACCAATATACATTAACAACTAATGATGAGGCTTGT 1000
Db 325186 TAAATATTTTAAATGCTATTTTAAATAATAATAATTAATTTTATTTTAAATTTT 325245
QY 1001 TTTTATTTGACATCAATTTGAGTTATCTTATGATGATCATGAAAAAATTTCCAACT 1060
Db 325246 ATTTATTTTAAATTTTAAATTAATTAATTTTATTTTAAATTAATTAATTTTAAATA 325305
QY 1061 GGTATATCTTTTACGAACACTATTCAGAACCAAGCTTAATCCCTACTTTTATGATGATG 1120

```

```

Db 325306 AATTAATACGATATTTTCTTTTGA-----TTTTTTAAATTAATTAATGATA 325357
QY 1121 ACATATTTTTCATATATATAGATGATGAAGATGATATTTTACGATGATTATTAAATTA 1180
Db 325358 ATTAATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 325417
QY 1181 AATTAATACGCTGATTTACGTCAAATGATGTAATAATTAATTAATTAATTAATGAT 1240
Db 325418 TAAGATGATTAATTTCTTTTATTTTATTCATTAATTTTAAATTAATTAATTAATTA 325477
QY 1241 AGTTAATATCAATACAAATACCTTAA 1267
Db 325478 TCGATTAATTAATTAATTTTATTTTCA 325504

```

RESULT 7

```

US-09-286-488-10
; Sequence 10, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 3701
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-286-488-10

```

```

Query Match 2.8%; Score 63.8; DB 9; Length 3701;
Best Local Similarity 43.8%; Pred. No. 0.024;
Matches 422; Conservative 0; Mismatches 532; Indels 9; Gaps 3;

```

```

QY 6 ATACTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 65
Db 2157 ATACACAACAANAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 2216
QY 66 AGAATCATTTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 125
Db 2217 TATTAATTTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2276
QY 126 TACTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185
Db 2277 TATCTTACTTGGGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2336
QY 186 GACTTTCGATCTTTTATGAGCGATTTTAACGATGTAAGGTAAGGTAAGGTAAGGTAAGG 245
Db 2337 ATCTTACTCTTTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2396
QY 246 AATATCAAAAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 305
Db 2397 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2456
QY 306 TTTACTTAAGACCATGCGTTATTTTGTATTAATTAATTAATTAATTAATTAATTAATTA 365
Db 2457 TGTATCTGTAATGAGCTTCTTATCTGTAATTAATTAATTAATTAATTAATTAATTAAT 2516
QY 366 ATCTGGAAGAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 425
Db 2517 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2576
QY 426 TCAAAATGTAATTTTGAAGAAAGCTGTTAAATCTTCAATCAATTTTGAATTTTAAATGTT 485
Db 2577 GAATTTGGAATTTTATGTAATTAATTAATTTAAAG---AAGTGAATTAATTAATTAATTA 2632

```

Query Match	2.8%	Score 63.8	DB 10	Length 3701
Best Local Similarity	43.8%	Pred. No. 0.024		
Matches 422	Conservative	0	Mismatches 532	Indels 9
			Gaps	3
QY	6	ATTACTAGATGCGATGCATATCATTAACAAACGACGTCGATGTTTATATCGGGATAT	65	
Db	2157	ATCACCAACAAATTAATTTTCACATCCCAATTAATTTGATTTTGGATT	2216	
QY	66	AGAAACATTAGCGGTACATATAAGTTAACGGACGAGAAAAAACCAACCAATTAATAAAACGT	125	

RESULT 9
 US-09-754-853A-2/c
 : Sequence 2, Application US/09754853A
 : Publication No. US20030005491A1
 : GENERAL INFORMATION:
 : APPLICANT: Haugse, Brian M.
 : APPLICANT: Parnell, Laurence D.
 : APPLICANT: Parsons, Jeremy D.
 : APPLICANT: Wang, Ming Li
 : TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With


```

; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45163)..(45164)..(45450)..(45509)..(46941)..(48763)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-2

```

```

Query Match          2.7%; Score 61.2; DB 9; Length 335913;
Best Local Similarity 43.3%; Pred. No. 0.1; 452; Indels 6; Gaps 1;
Matches 350; Conservative 0; Mismatches

```

```

QY 452 TTAATCTCAATCATTAGATTAAACAGTGTGTTAAATGGTTTAAATTAATTA 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99609 TTTATTTTCAATTTTAAAGTGTGAATATTTTATTTATTTATTTAATAA 99550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 TTGATACCTTATGAAGAACCAATACATCAATTCGACATTAGTAGAATTCCTGATG 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99549 ATATTTTAAATTTTACTCTTATAGTTTATTTAATTTTGTAGTATTTTATG 99490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 GTGGTTATTAAGAGATTCACACTTAACAGATTTTATTTATGACTTTTGTGAAG 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99489 ATTTTATTTAATAGTATATATTTTATTTTAAATTTTATTTATTTTAAAT 99430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 ATAAATGATGATAGTAGAAGCCCTAGACTATGCTGGAATGTTTGCAAAACCTCA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99429 GTTTATTTAAATTTTGTGATTTTAAATTTGAGATTTTAAATTTTAAATTTT 99370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 CACCTGACAACTTACATACATCATATGACGCTATTATATAGTATGCGCATATTC 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99369 TGTATATTACATATCGTATATAGCGCTCGTACAAATATATGACACTAAGATAT 99310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 ATTAATGATATATTTCCAAATTTTGACTATTAACAATTAACATTTTCAATTGAAATTA 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99309 TTTATTTTCAATTTTATATCAATGATTTTGTCAATTTTAAATTTTAAATTTT 99250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 TGGAACTCTTACTGATTAAGAAATGACAGCTTTTCAGTACACCAATATCAAGATA 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99249 TTAATTAATTTTACATGTTGTTATAGTAAATTTCTTACACAAATTAATTAAT 99190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 872 TTAATATCTTATACACTTATTCATTTCCATATGATTAATTTTATGACTATATTAAT 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99189 TTTCAATATTTAATGTTAATATATGCTACAA-----AATATTTAATCTTTTAA 99136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 932 CATCTATCGTGGTGTAAATATGTAATACACCAATATACATAAACAATTAATGATG 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99135 TTTTAGTTTAAATTTAAATTTAAAGTATTTAAAGAAACATATTTAAATATATTTA 99076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 992 AGCCTGTTTCTTATGACATTCAGTTATCTCTTATGATGATATCATGATAA 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99075 ATTTTATTTTATTTTCAAAATCAACTCACTTCAATATACATTTTGTGAATATTAATA 99016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1052 TTCCAACATGGTTACTTTAGACACATATTCAGAACCAACGTTAATCCCTACCTTTT 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99015 ATTTTAAATATACAAATATATTTAAATGAACTATACATAGTAAATTTAAATTTAT 98956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1112 TAGATATGACATTTATTTTCAATTTATATAGATTTGATTAAGATGTTTAAAGATG 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98955 TTAATTAATTTAATTAATTAATTTATTTATATTTTAAAGAAATTTGAATTAAGATAT 98896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1172 TATTAATTTAAATTAATTAATCACTGATATGCTCAATGATTTGTAATTAATTAATG 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98895 AAAAATATCAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 98836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 1232 ATAAATGATTCGTTAATATATACAA 1239
    ||| ||| ||| ||| |||
DB 98835 GATTTCAGATCAATAAACAATAA 98808
    ||| ||| ||| ||| |||

```

RESULT 10

```

US-09-754-853A-3/c
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763)..(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-3

```

```

Query Match          2.7%; Score 61.2; DB 9; Length 335913;
Best Local Similarity 43.3%; Pred. No. 0.1; 452; Indels 6; Gaps 1;
Matches 350; Conservative 0; Mismatches

```

```

QY 452 TTAATCTCAATCATTAGATTAAACAGTGTGTTAAATGGTTTAAATTAATTA 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99609 TTTATTTTCAATTTTAAAGTGTGAATATTTTATTTATTTTATTTAATAA 99550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 TTGATACCTTATGAAGAACCAATACATCAATTCGACATTAGTAGAATTCCTGATG 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99549 ATATTTTAAATTTTACTCTTATAGTTTATTTAATTTTGTAGTATTTTATG 99490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 GTGGTTATTAAGAGATTCACACTTAACAGATTTTATTTATGACTTTTGTGAAG 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99429 GTTTATTTAAATTTTGTGATTTTAAATTTGAGATTTTAAATTTTAAATTTT 99370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 CACCTGACAACTTACATACATCATATGACGCTATTATATAGTATGCGCATATTC 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99369 TGTATATTACATATCGTATATAGCGCTCGTACAAATATATGACACTAAGATAT 99310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 ATTAATGATATATTTCCAAATTTTGACTATTAACAATTAACATTTTCAATTGAAATTA 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99309 TTTATTTTCAATTTTATATCAATGATTTTGTCAATTTTAAATTTTAAATTTT 99250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 TGGAACTCTTACTGATTAAGAAATGACAGCTTTTCAGTACACCAATATCAAGATA 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99249 TTAATTAATTTTACATGTTGTTATAGTAAATTTCTTACACAAATTAATTAAT 99190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 872 TTAATATCTTATACACTTATTCATTTCCATATGATTAATTTTATGACTATATTAAT 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99189 TTTCAATATTTAATGTTAATATATGCTACAA-----AATATTTAATCTTTTAA 99136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 932 CATCTATCGTGGTGTAAATATGTAATACACCAATATACATAAACAATTAATGATG 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99135 TTTTAGTTTAAATTTAAATTTAAAGTATTTAAAGAAACATATTTAAATATATTTA 99076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 992 AGCCTGTTTCTTATGACATTCAGTTATCTCTTATGATGATATCATGATAA 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-09-938-842A-4551/c
Sequence 4551, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Krepes, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIORITY APPLICATION NUMBER: US 60/227,866
PRIORITY FILING DATE: 2000-08-24
PRIORITY APPLICATION NUMBER: US 60/264,647
PRIORITY FILING DATE: 2001-01-16
PRIORITY APPLICATION NUMBER: US 60/300,111
PRIORITY FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 4551
LENGTH: 2000
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-4551

```

Db 189 AAAAACAAATTTATATCTTTTCAA 164

	Query Match	2.5%	Score 57.6	DB 10	Length 431
	Best Local Similarity	48.2%	Pred. No. 0.2		
	Matches 162	Conservative	0	Mismatches 174	Indels 0
QY	313	AAAGACACATCCGTATTTTGATATATATTCACGCGAAATATATATTTAAATCTGCA	372		
Db	416	AAATTAATAAATAATTAATATATAAATAAATAATTAATTAATTAATTAATTAATTAATTA	357		
QY	373	GAAAGAAATGAAACACACATTTAAAAATGAAGAGCTCTATTTTAGCCAAAATCAAAAT	432		
Db	356	ATTAATAAATTTAAAAATAAATAAATAAATAAATAAATAAATTAATTAATTAATTAATTA	297		
QY	433	GTATTTTGAAGAAACGTAAATCTCATCATTAATTAGATTTTAACAATGTTTAAAT	492		
Db	296	AAACAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT	237		
QY	493	GGTTTAAATTTAAATATTTGATTACTTTTGAAGAACCAATACATCAATTTGCACATTA	552		
Db	236	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAATAAATAAATAAATAAATAA	177		
QY	553	GGTAGAATAATTACTTGATGGTGGTTATTTTAACAGAACTCACACTTAAACAGATTTTAAT	612		
Db	176	AAATTAATAAATAAATAA	117		
QY	613	TATACATTTTGGATAAAGATAAAGATATGAGATAT	648		
Db	116	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAATAAATAAAT	81		

RESULT 13
 US-09-790-988-1/c
 : Sequence 1, Application US/09750988
 : Patent No. US20020127687A1
 : GENERAL INFORMATION:
 : APPLICANT: SHIGENOBU, SHUJI
 : APPLICANT: MATSUNBE, HIDEKI
 : APPLICANT: HATTORI, MASAHIRA
 : APPLICANT: SASAKI, YOSHITAKI
 : TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
 : FILE REFERENCE: 081356/0159
 : CURRENT APPLICATION NUMBER: 05/09790, 988
 : CURRENT FILING DATE: 2001-02-23
 : PRIOR APPLICATION NUMBER: JP2000-107160
 : PRIOR FILING DATE: 2000-04-07
 : NUMBER OF SEQ ID NOS: 7
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 1
 : LENGTH: 640681

Query Match:	2.4%	Score 56:	DB 9:	Length 2000:
Best Local Similarity	48.9%	Pred. NO. 0.42:		
Matches 178:	Conservative	0:	Mismatches 185:	Indels 1:
Gaps				
Qy 912	TTTTTATGACATATATTAAATTCATTCCATCGTGGTGTAAATATGATATACACCAATA	971		
Db 1979	TATTTATTAGAAAATATGACATCCATTGGTGTATATATATATATGAAATGGAAGCA	1920		
Qy 972	CATTAACAAACTATATGATGAGCCCTGTTTTCTATTTGACATTCATTCGAGTATCCTTA	1031		
Db 1919	GACACACATTCACAAAAAAGTAAATGGTTTCCTTTTTTTCCTGTTGAGTGTGA	1860		
Qy 1032	TGATGTATATCAGAAAAAATTCACATATGTTATATCTTTACAGACATTCAGAAC	1091		
Db 1859	CTTGACCTTAAATAAAAGTGTGTACATCTTTACCTGACAACTCTATATATATAT	1800		
Qy 1092	AACGTATATCCCTACTTTTATAGATGATGACAAATATTTTCAATATATAGATGTATA	1151		
Db 1799	GTTTTCCTTTTTTCTTGAAAGAACTTTATATATATATACAT-TATATGGTTATTA	1741		
Qy 1152	AGATGTATTAACGATGATTTATTTATTAATTAATTAATCAGCTGATATAGCTCAATGAT	1211		
Db 1740	TTTTTTTTTTGGCAACTTTTCTTANTTCATTTATATCTCCTTATAGGATTTTTTT	1681		
Qy 1212	TGTAAATACTATATATATGATATATGATACGTTATATATTCATTCACAAATCAATTAGACT	1271		
Db 1680	TGAATAATTCAAATTTGGTAAATATAAAATAAAATAATCTCTTAAAAAATAATATCTAT	1621		
Qy 1272	GATT 1275			
Db 1620	AATT 1617			

Search completed: January 8, 2003, 13:57:40
Job time : 5443 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 01:09:30 ; Search time 2138 Seconds
(without alignments)
17316.596 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 2286

Sequence: 1 atggatctactagatgcacgtcttaaaaggaagcaactgttaa 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8095743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hiv:*
20: em_gss_pln:*
21: em_gss_vrl:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111.6	4.9	1101	17	AL069706 Drosophila
2	104.2	4.6	1101	17	CNS000EVL
3	89.6	3.9	994	17	CNS004NOI
4	89.2	3.9	1101	17	CNS0021J
5	89.2	3.9	1101	17	CNS017KE
6	88.8	3.9	1007	17	CNS06X9S

Result No.	Score	Query Match	Length	DB ID	Description
7	88.2	3.9	1101	17	CNS001FB
8	88	3.8	1101	17	CNS0039G
9	87.6	3.8	1101	17	CNS0039G
10	87.4	3.8	1021	17	CNS014DY
11	83.4	3.6	1225	17	CNS016D
12	83.2	3.6	1101	17	CNS0021J
13	82.6	3.6	1101	17	CNS003BD
14	81.8	3.6	1101	17	CNS00E07
15	81.6	3.6	820	17	B11728
16	81.4	3.6	829	17	CNS03LUB
17	81	3.5	1092	17	CNS020K7
18	81	3.5	1101	17	CNS00E07
19	80.6	3.5	1161	17	CNS073Y8
20	80.4	3.5	1190	17	CNS020N7
21	80.2	3.5	887	12	BF274559
22	80.2	3.5	945	17	CNS04DOK
23	80	3.5	1146	17	CNS021G2
24	79.8	3.5	781	17	CNS009D0
25	79.4	3.5	1292	13	BM463105
26	79	3.5	734	17	CNS010MP
27	79	3.5	987	17	CNS014PQ
28	78.8	3.4	668	9	AL514901
29	78.8	3.4	876	17	CNS009G1
30	78.8	3.4	1152	12	BG309087
31	78.6	3.4	1161	17	CNS073Y8
32	78.4	3.4	1201	17	CNS0167M
33	78	3.4	1248	17	B11336
34	77.4	3.4	661	17	CNS028VY
35	77.4	3.4	928	17	CNS00DKY
36	77.4	3.4	1169	17	CNS06KHQ
37	77	3.4	930	17	BH135412
38	76.8	3.4	1098	12	BG845178
39	76.8	3.4	1225	17	CNS016D
40	76.6	3.4	1101	17	CNS00B01
41	76.4	3.3	1101	17	CNS003BB
42	76.2	3.3	1027	17	CNS02W50
43	75.8	3.3	1190	17	CNS020N7
44	75.4	3.3	661	17	CNS020V7
45	75.4	3.3	1187	17	B11102

ALIGNMENTS

RESULT 1
CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
LOCUS BACR29823 of RPI1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLES Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osogawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI1-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR29B23"

/clone_lib="RPCT-98"

/note="end : 77"

BASE COUNT

419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match

4.9% Score 111.6; DB 17; Length 1101;

Best Local Similarity 38.7%; Pred. No. 1.2e-09;

Matches 253; Conservative 119; Mismatches 274; Indels 8; Gaps 4;

337 AATATACAGCGCAAAATATATTTAAATCGCAGAGAAATGACACACATTAATA 396

451 MATTCGATCTTMMMMMMMAATWMAAATAATTTATWATAAATAAAMMMWATTT 510

397 ATGAAAGAGGCTACTATTTAGCCAAATCAAAATGTAATTTAGAAAAAGCTTAA 456

511 TTTMMWMTWATTTTWWMTWMTWMAAAAAAAATATTAATWMAATWATTA 570

457 TCTTCATCATTTAGATTTAACAATGTTTAAATGTTTAAATTTATATTTGAT 516

571 ATTAA--AAWTTATATTAATTAATWMTWMTWMTWMTWMTWMTWMTWMTW 627

517 AACTTTATGAAACCAATACATTCATTCGACATTAGTAGTAATTTACTGCTGT 576

628 AATTTTATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 687

577 TATTTACAGATTCACACATTAATAACAGAT--TTAATTAACGATTTTGAATA 634

688 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 747

635 ATGATATGATGATAGTGAAGCTATGACTATGCTGT--GAATGTTTGCACAACT 693

748 AATATATATATWMTATATATWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 807

694 CCGACCACTTCATCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 753

808 AATATATATWMTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 867

754 TATAGATATATTTCCAAATTTTGACTATTAACAAATTAACATTTTCATTAAT 813

868 WATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 927

814 GATCTTACTGATATATGAAAGACAGTTTCATTCATTCATTCATTCATTCAT 873

928 AATATATATWMTATATATTTAATAATWMTATWMTATWMTATWMTATWMTAT 987

874 AATATATCTATACATTTATCATTTCCATGATA--TGAATTTTGTGCTATTTAT 931

988 TATATATWMTATATATATTTAATAATWMTATTTTAAATAAATAAATAAATA 1047

932 CATTCATCTGCTGCTGTTTAAATATGTAATACCAATTAACATTAACATTA 985

1048 AATATATWMTATATATTTAATAATWMTATWMTAATAAATAAATAAATA 1101

RESULT 2

CNS00EVL/c

LOCUS

DEFINITION

ACCESSION

VERSION

AL069706.1

GI:4949849

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNML

COMMENT

GSS.

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota:

Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:

Ephydroidea: Drosophilidae: Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCT-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2: cn bw sp, the same strain used for the BDGP's

pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR29B23"

/clone_lib="RPCT-98"

/note="end : 77"

BASE COUNT

419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match

4.6% Score 104.2; DB 17; Length 1101;

Best Local Similarity 35.7%; Pred. No. 2.2e-09;

Matches 225; Conservative 112; Mismatches 291; Indels 3; Gaps 1;

327 TTATTTGATATATTTTACACCGGAATATATTTAAATCTGCAGAGAAATGACA 386

1090 TTWTTTWTWMTATATTAATTAATTAATTAATTAATTAATTAATTAATTA 1051

387 CACATTTAAATGAAGAGGCTACTATTTAGCCAAATCAAAATGATTTAGAAAA 446

1030 TTTWAAAATATATWTTTAAATATATWMTATWMTATWMTATWMTATWMT 971

447 ACGTGAATCTTCATCAATTTAGATTTAACAAGTTTAAATGCTTTAAATTTAA 506

970 WATWATWMTATWMTATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 911

507 TATTTTGTATACCTTATGAAGAACAATACATTCATTCGACATTTAGTAAGAAT 566

910 TWTWTTTATATWMTWMAAAMWMTATWTTTWTWMTWMTATWMTATWMTATWMT 851

567 TGATGCTGGTATTTTACAGATTCACACTTAAACAGATTTTAA--TATACGATTT 623

850 TWTWATTTWMTATATWMTATWMTATWMTATWMTATWMTATWMTATWMTATWMT 791

624 TGATTAAGATTAATGATATGATAGTGAAGCTATGACATCTGCTGAATGTTTGC 683

790 ATWTTATWMTWMTATWMTATWMTATWMTATWMTATWMTATWMTATWMTATWMT 731

684 AAAACTCACACCTGCAACACTTACATCATTAATGACGCTATTAATAGTATGTG 743

730 TTTAATWMTATWMTATWMTATWMTATWMTATWMTATWMTATWMTATWMTAT 671

744 CCATATTCATTAATAGATATTAATTTCCAAATTTGACTATTAACAAATTAACATTT 803

670 AWWMTATWMTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 611

OY		804	GAAITTTTNGCAATCTTCGTAAGAAGTGACGGTTCGTACTGACCGAACACAATA	863
Dd		610	TATTATTATANMAWMTTATTATATATATAMWTTMATTTAATTTTATATWATTMWTITAAA	551
OY		864	TCACATCATTTAAAAATCTTTATACACATTATCATTTCATGATGAATGAAATTTTATGACCTA	923
Dd		550	TTATTTTTTTTTTTTTTTTAMAWMTTAAAAAATAATMMWWMAAAATATMMWTTTWTATTW	491
OY		924	TATTATCATCTCTACGTGC GTTTAAAT	954
Dd		490	ATAATTTTWTMTWMWMTTMMKKKMKKMAD	460
RESULT 3 CONS4NOJ LOCUS				
DEFINITION Tetraodon nigroviridis genome survey, sequence T7 end of clone 122P04 of library G from Tetraodon nigroviridis , genomic survey sequence.				
ACCESSION ALZ98972.1 GI:8037822				
VERSION ALZ98972.1				
KEYWORDS GSS; genome survey sequence.				
SOURCE Tetraodon nigroviridis . Tetraodon nigroviridis . Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; NeoTeleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontidae; Tetraodon. 1 (bases 1 to 994) Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Winkler,P., Brottier,P., Queller,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence Unpublished				
REFERENCE 2 (bases 1 to 994)				
AUTHORS Roest-Crollius,H., Jalllon,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and Weissenbach,J. TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Unpublished				
COMMENT 3 (bases 1 to 994) Genoscope. Direct Submission Submitted (12-APR-2000) This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.				
FEATURES Source Location/Qualifiers 1..994 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /cdate="122P04" /cloned_lib="G" /note="Genoscope sequence ID : CORBGI22DH02LE1-end : T7"				
BASE COUNT 543 a 49 c 36 g 194 t 172 others				
ORIGIN				
Query Match 3.9% ; Score 89.6; DB 17; Length 994;				
Best Local Similarity 43.9%; Pred.No.7.2e+06;				
Matches 281; Conservative 27; Mismatches 325; Indels 7; Gaps 2;				
OY		350	AAATATATTATTAAATATCTGCACAGAAATGACACACATTAAATAATGAAAGCGCTA	409
Dd		326	AATNMNNAANNANA	385
OY		410	CATTTTGCCAAATCAAATGTAATTTTAGAAAACGCTTAATCTTCAATCAATT	469
Dd		386	AAAAAAAAATTANANNAAAAAAAAAATAAANAATTTTATATATWTAATWANMATTNANTA	445
OY		470	TAGATTAAACAATGTTTTTAAATGGTTTTAAATTTTAATATATATGATACTTATGAAA	529

Db 446 ATATATAAAAAAATTATTTAAATNAATTAATTANTATNTATNTATTTATTTATNTATNTATAAWA 505

QY 530 CCAATTCACATCAATTGCCAACATTTAGCTAAGAATAATTACTTGATGGTGTGTTATTTAACAGAT 589

Db 506 TAAAATATATATATATANAANNTAAAAAATAATWTATTTAATNMTAATATATATXT 565

QY 590 CACAACCTTAAACAGATTTTATTTATACGATTTTGATTAAGATATCATATGAATGATA 649

Db 566 AATTAATATAAAAAATAAATNATNMTAATATATAAAAAAATATATAAAAATWTTTATAANAW 625

QY 650 GTGAGCCATATGACATATCGTGTGAATGTGTTTTGCAAAACCTCACACCCTGACACACTTACAT 709

Db 626 ATTAATATATATATATW-----TTTATATATATANAATNTATNTATATATATTTATTTTAW 680

QY 710 ACATTCATATATGACGTGATTTATTTAGTAGTATGCGCATATTCATTATNGCATATATTTTC 769

Db 681 ATTTPAAAAMNNAATATAMNNMAAATAAAMAAATAAAAAAAAAAAAAAAAAAAAAA 740

QY 770 CAATTTTGACATATACAAATTAACATTTTCATGCAATATATNAGCAATCTACTGATATA 829

Db 741 AAATTTTATATATATANAAMAAAAATATATMAAAAAAATAATATATATMAATMAAAAA 800

QY 830 ATGAATGACACAGTTTCAGTTACTCAACCAATATCAGATTTTAAATATCTTATACAC 889

Db 801 AWAATAAAAAAATAATTAATTTTANTMTAAAAAATAAATAATATATRTTATATATAM 860

QY 890 ATTATCATTTCCATGATATGATTTTATNTAGCTATATTAATNCATCTATCTGTGCTT 949

Db 861 AAAATNTTATTAATAAAMWAATAATANATTTA-ATATATAAAMATTAATTTTGTCTA 918

QY 950 TAAATATGTATACACCAAAATACATMAACAACATTAATGA 989

Db 919 AAMATATTTTAAAAAATAATTAATTTAATTTGATTTAA 958

RESULT 4
CNS0021J 1101 bp DNA linear GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TEI3 end of BAC # BAHR05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

LOCUS CNS0021J

VERSION AI061936

KEYWORDS GI:4940214

SOURCE GSS.

ORGANISM Drosophila melanogaster.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
<http://www.genoscope.cns.fr>)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuoyo Goessawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PL and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
1..1101

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR0511"
/clone_lib="RPCT-98"
/note="end : 1E13"

```

Query Match	3.9%;	Score 89.2;	DB 17;	Length 1101;
Best Local Similarity	39.9%;	Pred. No. 8.3e-06;		
Matches 304;	Conservative 58;	Mismatches 398;	Indels 2;	Gaps 1;

malices 304; conservative 38; mismatches 390; liners 4; gaps 17

malices 304; conservative 38; mismatches 390; liners 4; gaps 17

Accession	Position	Sequence	Gene
QY	226	AAAAGACGGTATCAATTCACAAAACAGAGTATATCAAGATTCGACATCTG	285
Db	259	AAAG	318
QY	286	AATAATACATATATCTTTTACTTAAGACACCATGCTTATTTGATATATACA	345
Db	319	AA	378
QY	346	CGCGAATATATATTTAATAATCTGCAGAGAAATGACACACATTAATAATGAAGAG	405
Db	379	AA	438
QY	406	GCTACTATTTTACCCAAAAATCAAAATGTAATTTAGAAAAACGTAAATCTTCATC	465
Db	439	AA	498
QY	466	AATTGATTTTACATCTTTTAAATGCTTTAAATTTAATATATATGTAATCTTAA	525
Db	499	AAAAAAAAAAAAAAAAAAAAAAAAATATATATTTTATTTTAAATTTTATTTT	558
QY	526	AAACCAATTCATCATATTCGACATTTAGTAAAGAAATTTCTGATGCTTTTACAC	585
Db	559	WTATATTTATTTTAAWAAATAATTAATAAATAAAWTTTATTTTAAAT	616
QY	586	GAAATCACACTTAAACAGATTTTATATACAGATTTTGATTAAGATATATGATAT	645
Db	617	AAAAAAAAATTTTAAAWTTTATTTTATTTTATTAATTAATAAATAAAATTAATTT	676
QY	646	GATAGTAAACCTATGATCTGTCGAAATGTTTGGAAAACACACCTGACCACTT	705
Db	677	ATWATWAAATWAAATTTTAAAAWTTTATTTTATTTTAAAAWTTTATTT	736
QY	706	ACATACCTATATATGACGTGATATATAGTATAGTCCATATTCATATAGCATATA	765
Db	737	TAAAAWTTTAAWTTTATTTTATTTTATTTTATTTTATTTTATTTT	796
QY	766	TTTCCAAATTTTACATTTACAAATTTACATTTTATTTATTTATTTATTTT	825
Db	797	TTTTAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	856
QY	826	AATATGAAATGACAGCTTTTCACTTCACTCAACAAATTCAGATATTTAAATATCT	885
Db	857	TTTAAATTTTATTTTATTTTAAATAATTTTAAATATTTTAAATTTTAA	916
QY	886	ACATATATTCATTTCCATGATATGATTTTTCGCTATTTAAATCATCTATCGG	945
Db	917	AAATGGAATGKKGKGTARAKKGGKGTATARATDSTAGARATPAKKTATKATK	976
QY	946	GATTAAATATGTATACACCAATATACATAAACAACATAAT	987
Db	977	TAAAAADKAKGAAATATWGAAARCATGAATAAGTGDAT	1018

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

VERSION ALL08152.1 GI:5628456

KEYWORDS
GSS.
Diosophila melanogaster.
SOURCE
Drosophila melanogaster

RESULT 12	CNS0021J/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
			CNS0021J					
			Drosophila melanogaster genome survey sequence Tetr3 end of BAC #					
			BACR03N11 of RPCT-98 library from Drosophila melanogaster (fruit					
			fly), genomic survey sequence.					
			AI061936					
			AI061936..1 GI:4940214					
			GSS.					
			Drosophila melanogaster.					
			Drosophila melanogaster					
			Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prelygota;					
			Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;					
			Ephydroidea; Drosophilidae; Drosophila.					
			1 (bases 1 to 1101)					
			Genoscope.					
			Direct Submission					
			Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage					

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RBC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the library P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

Query Match	3.6%	Score 83.2	DB 17	Length 1101
Best Local Similarity	39.6%	Pred. No. 9e-05		
Matches 261	Conservative 59	Mismatches 335	Indels 4	Gaps 1

[illegible]

RESULT 13				
CNS003BD				
LOCUS				
DEFINITION	CNS003BD	1101 bp	DNA	linear
	Drosophila melanogaster genome survey sequence TET3 end of BAC #			GSS 03-JUN-1998
	BAC008K08 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL064091			
VERSION	AL064091.1	GI:4941847		
KEYWORDS				
SOURCE	GSS.			
ORGANISM	Drosophila melanogaster.			
	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
	Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;			
	Ephyrididae; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 1101)			

Db 802 WMAATWTTATAMNATTTTTTTTANAAWTTWATWATATTTTAAWTTTAAATA 743

QY 707 CATACATTCGATATGACGCGATATTAATAGTATGCGCATATTCATTTAGCATATAT 765

Db 742 AAAAATTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 690

QY 767 TTCCAAATTTTGACTATACCAATTAACATTTTCATTTGATATTTATGATCTTACTGA 826

Db 689 AATTAATWTTTAAATWTTAAATWTTAAATWTTTAAATWTTAAATWTTAAATTTTA 630

QY 827 ATATGAAATGACAGCTTTTCACTCTACACCAATATCAGATATTAATATCTTA 886

Db 629 WAAWTTATATWTTTATTTTTCWCTWATATATTAATWAAATTAATTAATTTTATAT 570

QY 887 CACATTAATCATTCATTCATGATGATGATTTTATGACTATATTTAAATCATCTGATG 946

Db 569 WMAATTAATAAAATWTTTATTTATTAAMWATTTATTWATWTTGATTTTATTTATTT 510

QY 947 GTTAAATATGATTAACACCAATTAACATTAACAACTAATGATGAGCTTTGTTCTA 1006

Db 509 TTTTAAATTTTATTTATATWATTTATTAATWTTATATTTTATTTATCTTCMAA 450

QY 1007 TT 1008

Db 449 WT 448

RESULT 15

B11728 820 bp DNA linear GSS 14-MAY-1997

LOCUS B11728/c

DEFINITION T8D18-Sp6.1 TAMU Arabidopsis thaliana genomic clone T8D18, DNA

SEQUENCE

Accession B11728

Version B11728.1 GI:2092850

Keywords GSS.

Source thale cress.

Organism Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 820)

AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.

TITLE BAC End Sequences at ATGC

JOURNAL Unpublished (1997)

COMMENT Other_GSSs: T8D18-T7, T8D18-Sp6

Contact: Ecker J.

Arabidopsis Thaliana Genome Center

University of Pennsylvania

Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104

Tel: 215-898-9384

Fax: 215-898-8780

Email: jecker@atgenom.bio.upenn.edu

Seq primer: Sp6

Class: BAC ends

High quality sequence start: 411

High quality sequence stop: 423.

FEATURES

Source

1. 820

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="T8D18"

/clone_1lb="TAMU"

/sex="hermaphrodite"

/note="vector: BelobACII; Site_1: HindIII; Site_2: HindIII

; Produced by Rod Wing

BASE COUNT 248 a 18 c 33 g 360 t 161 others

ORIGIN

Query Match 3.6%; Score 81.6; DB 17; Length 820;

Best Local Similarity 41.7%; Pred. No. 0.00017;

Matches 288; Conservative 0; Mismatches 394; Indels 9; Gaps 3;

QY 329 ATTTGATTAATTTTACACCGCAAAATATTTATTTAAATCTGCAGAGAAATGACACA 388

Db 802 ATTTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 743

QY 389 CATTAATAATGAAGAGCTACTATTTTACCAAAATCAAAATGTAATTTTGAAGAAC 448

Db 742 ATAAAAAATATTTATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 683

QY 449 GTTTAAATCTTCAATTTTATGATTTACATGTTTAAATGTTTAAATTTTAAATTA 508

Db 682 AATAAATAATTTTAAATTTATTTATTTAAATTTTAAATTTTAAATTTTAAATTA 623

QY 509 TATGATTAAGCTTTTGAAGAACCAATACATTCATTCGCAATGATAGGTAAGATCTG 568

Db 622 AANNNTT-TAAATATTAANNNAAANNNAAANNNAAANNNAAATTTAANNNAAATTN 564

QY 569 ATGTTGTTATTTTAAACAGATCAACCTTAAACAGATTTTATATAGATTTTGATA 628

Db 563 ATTAATAATTTTATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTA 508

QY 629 AAGATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 688

Db 507 AAAAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTA 448

QY 689 TCACACCTGACACACTTACATCATTCATTAATGAGTGATTTATAGTATGCGCAT 748

Db 447 TAATTTTNNNTTNNNTTNNNNAAANNNAAANNNAAANNNAAATTTTAAATTTAAAT 388

QY 749 TGCATTAATGATATATTTTCCAAATTTGACTATTAACAAATTTACATTTTCAATGA 808

Db 387 NAAATNNNTTNNNTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 332

QY 809 TTATGATCTTACTGATGATTAATGAATGACAGCTTTGATGATGATGATGATGATG 868

Db 331 ATTTAATTTTATTTTAAANNNNNAAATTTATTTTNTTTTAAATTTTAAATTTAA 272

QY 869 ATATTAATATCTTATACACATTTATCAATTCATGATGATGATGATGATGATGATG 928

Db 271 TTAATAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 212

QY 929 AATCATCTATGCTGCTGCTGCTTAAATATGATATACACCAATATACAAATGATG 988

Db 211 ATTTTAAATTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNN 152

QY 989 ATGAGCTTGTATTTCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1019

Db 151 NNAATWTTNNNNNNNGGNNCCNNNNC 121

Search completed: January 8, 2003, 11:19:44

Job time : 2153 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 50.5 Seconds
(without alignments)
12063.799 Million cell updates/sec

Title: US-09-727-892A-2
Perfect score: 4070
Sequence: 1 atcgagactactagatgcattgcatt.....taaaaaaggaacacctgttaa 2286

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_n2p_model -DEV=xlh
-Q=C9n2_1/USPTO.spool/US09727892/r/natc_06012003_151200_9278/app_query.fasta.1.2439
-DB=A.Geneseq_101002 -QPMT=fastan -SOFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US09727892.gcch.1.1.54.tunat.06012003.151200.9278 -NCPU=6 -ICPU=3
-NO.XLPXY -NO.MMAP -LARGEQUERRY -NEG.SCORES=0 -WAIT -LONGLOG DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: A.Geneseq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4069	100.0	761	21	AA16529
2	351	8.6	73	21	AA16547
3	257	6.5	50	21	AA16563
4	242	5.9	121	21	AA18241
5	237.5	5.8	3973	21	AA18253
6	230	5.7	1802	21	AA18217
7	229.5	5.6	1182	21	AA18288
8	225	5.7	42	21	AA16580
9	220	5.4	573	21	AA16603
10	218.5	5.4	872	23	AB17619
11	218	5.4	40	21	AA16585
12	218	5.4	767	23	AB17616
13	218	5.4	1817	21	AA18301
14	212	5.3	42	21	AA16581
15	211	5.2	1417	20	AA13947
16	208	5.1	1617	21	AA18255
17	206	5.1	1811	20	AA13948
18	205	5.0	1188	21	AA18183
19	202.5	5.0	2485	21	AA18172
20	202	5.0	1149	23	AB17620
21	201.5	5.0	2206	21	AA18254
22	199	4.9	1381	23	AB18254
23	198	4.9	797	21	AA18179
24	198	4.9	4643	22	AB17609
25	196	4.8	2013	21	AA18265
26	194	4.8	1123	21	AA15827
27	193	4.7	1516	21	AA18195
28	192.5	4.7	609	23	AB17640
29	191.5	4.7	1247	21	AA18215
30	191.5	4.7	1346	21	AA18236
31	190	4.7	2496	21	AA18222
32	189.5	4.7	2639	21	AA18198
33	188	4.7	36	21	AA18592
34	186	4.6	2380	21	AA18315
35	185	4.5	2184	20	AA13946
36	184.5	4.5	2184	22	AA18245
37	183.5	4.5	2710	17	AA195016
38	183.5	4.5	2710	19	AA18247
39	183	4.5	1245	21	AA18244
40	183	4.5	3135	15	AA18244
41	183	4.5	3135	21	AA18244
42	181.5	4.5	944	23	AB139507
43	181	4.4	1308	21	AA18167
44	180.5	4.4	1193	21	AA18306
45	178.5	4.4	507	21	AA18173

ALIGNMENTS

RESULT 1
AA16529
ID AA16529 standard; Protein: 761 AA.
XX
AC AA16529;
XX
XX 27-OCT-2000 (first entry)
DE Bacteriophage 44AHD protein sequence 44AHDORF001.
XX
XX Bacteriophage: antimicrobial; genome; identification; antibacterial;
KM bacterial growth inhibition; bacterial infection.
XX
XX Bacteriophage 44AHD.
OS
XX
XX WO200032825-A2.
XX
XX 08-JUN-2000.
XX


```

Db 501 ValLeuAnSngLyLeuTyGlyLePrCoLaLeuArSgSerHisPheAsnLeuPheArgLeu 520
QY 1561 GATGATTAACAATGACATATATCATATACGGTTACAAAACAGTGAACATAATA 1620
Db 521 AspAspAsnAnSngLyLeuTyArSniLeLeAsnGlyTyLySAsnThrGluArSniLe 540
QY 1621 TTATTCCTACATTTGTCACATCACGTTTCATTGATATAACTATTGGTTCCTTCCAATAC 1680
Db 541 LeuPheSerThrPheValThrSerArSgSerLeuTyArSniLeuValProPheGlnTy 560
QY 1681 TTAACGAAAGTGAATGACGACATTTTATTCGCTAGTACGATGATGATTTATATG 1740
Db 561 LeuThrGluSerGluIleAspAspAsnPheLeuTyArSniPheArgSerLeuTyMet 580
QY 1741 AAATCCGTTGTTAAACCCCTTATGACCCAGTTTATTCGACCGATAGCCCTAGGTAA 1800
Db 581 LysSerValValLysProLeuLeuAsnProSerLeuPheAspProIleAlaLeuGlyLys 600
QY 1801 TGGGATATTGAAAGACATAGATAGATGATGTTTGACTGATCATTAAGAAATATGCA 1860
Db 601 TrpAspIleGluAsnGluGlnIleAspLysMetPheValLeuAsnHisLysLysTyAla 620
QY 1861 TATGAGTGAATGGAAGATTAATAATTGCTCTGCTGCTATACCGAAACGCTTTGAT 1920
Db 621 TyrgluValAsnGlyLysIleLysIleAlaSerIleGlyLeuProLysAsnAlaPheAsp 640
QY 1921 ACAAGCTCGATTTTGAACCTTTGACGTGACATATTCCTTGAACGCTCATTTATGAA 1980
Db 641 ThrSerValAspPheGlnThrPheValArgGlnIlePheAspGlyAlaIleIleGlu 660
QY 1981 AACAATAAAGTATCTATATAGCAAGTACAAATGATGATATCCCTTAAACTGAA 2040
Db 661 AsnAsnLysSerIleTyArSniGlnIleGlnIleThrIleSerIleTyProSerLysThrGlu 680
QY 2041 ATGTGATGCTATGATGATGATGATATTTTACTGATGACCTAATATGAAACGTGAA 2100
Db 681 IleValCysGlyAsnValTyArSniLysPheThrAspGluLeuAsnMetLysArgGlu 700
QY 2101 TTTATATTAAGAAGCGGTAGAGAAATTTGACCACTGATGATGATGATGATTTAT 2160
Db 701 PheIleLeuLysAspAlaArgGluAsnPheAspHisSerGlnPheAspAspIleLeuTy 720
QY 2161 ATTGAAGTGAATGCGTTCATTTTCACTTAACGACTTATTTTCAGTTGAACGTTCAGTA 2220
Db 721 IleGluSerAspIleGlySerPheSerLeuAsnAspLeuPheProValGluArgSerVal 740
QY 2221 CATACCAATCTGATTTGCATATATTAACGTGACATGATGAAATTAAGGCAAC 2280
Db 741 HisAsnLysSerAspLeuHisIleLeuLysArgGlnHisAspGlnIleLysGlyAsn 760
QY 2281 TGT 2283
Db 761 Cys 761

```

RESULT 2

AA16547 ID AAB16547 standard; Protein: 73 AA.

XX AAB16547;

XX 27-OCT-2000 (first entry)

XX Bacteriophage 44AHD protein sequence 44HDDRF017.

XX Bacteriophage; antimicrobial; genome; identification; antibacterial;

XX Bacterial growth inhibition; bacterial infection.

XX Bacteriophage 44AHD.

XX WO200032825-A2.

XX 08-JUN-2000.

```

XX PF 03-DEC-1999; 99MO-IB02040.
XX PR 03-DEC-1998; 98US-0110992.
XX PR 03-JUN-1999; 99US-0326144.
XX PR 28-SEP-1999; 99US-0407804.
XX PR 30-SEP-1999; 99US-0157218.
XX PR 01-DEC-1999; 99US-0168777.
XX PR 02-DEC-1999; 99US-0454252.
XX PA (PHAG-) PHAGETECH INC.
XX PI Pelletier J, Gros P, Dubow M;
XX DR WPI; 2000-412361/35.
XX DR N-PSDB; AAA69032.
XX PT Identifying a bacteriophage coding region for treating bacterial
XX PT infections comprises identifying a nucleic acid encoding a product that
XX PT inhibits bacteria when a bacteriophage infects a bacterium
XX PS Example 9; Page 277; 456pp; English.
XX CC The present invention describes a method for identifying a bacteriophage
XX CC coding region encoding a product active on an essential bacterial
XX CC target. The method comprises identifying a nucleic acid sequence encoding
XX CC a gene product that provides a bacteria-inhibiting function when an
XX CC uncharacterised bacteriophage infects a pathogenic bacterium. The
XX CC compound active on a target of a bacteriophage inhibitor protein in a
XX CC bacteria is used to treat or prevent a bacterial infection in an animal.
XX CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
XX CC nucleotide and protein sequences which are used in the exemplification of
XX CC the present invention.
XX SQ Sequence 73 AA:
XX
XX Alignment Scores:
XX Pred. No.: 7.07e-24 Length: 73
XX Score: 351.00 Matches: 73
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 8.62% Indels: 0
XX DB: Gaps: 21
XX
XX US-09-727-892a-2 (1-2286) x AAB16547 (1-73)
XX
XX QY 1671 ATGGAAGATTAATAATGCTTCGCTATACGAAACGCTTGATACAAAGCTGC 1930
XX Db 1 MetGluArgLeuLysLeuLeuLeuValTyArGlyThrProLeuIleGlnAlaSer 20
XX
XX QY 1931 ATTTGAACCTTTGACGTGAACATTTCTTGACGCTGCCATTATGAAACATAAAA 1990
XX Db 21 IleLeuLysProLeuTyValAlaAsnSerLeuThrValProLeuLysThrIleLys 40
XX
XX QY 1991 GTATCTATATGACAGGATGCAATATCGATATATCCGCTTAAACGGAATGTATGTC 2050
XX Db 41 ValSerIleMetSerLysValGlnTyArGlyThrLeuArgLeuLysLeuLysLeuTyVal 60
XX
XX QY 2051 GTAATGATATGATGATATTTTACTGATGATGATGATGATGATGATGATGATGAT 2089
XX Db 61 ValMetTyMetMetAsnIleLeuLeuMetAsnLeuLeu 73
XX
XX RESULT 3
XX AAB16563 ID AAB16563 standard; Protein: 50 AA.
XX AC AAB16563;
XX XX 27-OCT-2000 (first entry)
XX DE Bacteriophage 44AHD protein sequence 44HDDRF031.
XX XX Bacteriophage; antimicrobial; genome; identification; antibacterial;

```


Db 105 TyrIleSerLeuCySAspIleIleGlnSerValIleLysIlePheAspGluLeuAspLysThr 124
QY 253 AAAACAGATATTATCATGATTGCACATRACTGTATATAATACGATTAATCATTTTACTT 312
Db 125 PheThrAspTyrAsnPhetYrIleGluValLysAsn---IleAspLysAsnValIleAsn 143
QY 313 AAAGACACCATGCGTTATTTTGATTAATAT-----TACAGCGCAAAATATATA 359
Db 144 LysIleAsnGluIleTyrPheLysAsnLysAspIleThrPheHisIleAspGluIleLeu 163
QY 360 TTTAAATCTGCAGAGAAATGACACACATTAATAATGAAAGAGC----- 407
Db 164 GlyLysIleCySAsnLysIleMetSerTyrIleHisGluMetAsnGluLysAsnGluIle 183
QY 408 -----TACTATTTTAGCCAAATAATCAAAATGTAATTTTACAAAAACGTGTAATC 458
Db 184 HisPheLeuIleTyrPhePheArgIleTyrAsnLysAsnSp-----LysAsnLeuIleLeu 201
QY 459 TTCATCATTTAGATTATTAACAATGTTTAAATGCT---TTAAATTTAT----- 507
Db 202 PheTyrAsnTyrTyrPheAsnTyrValPhe-AspHisMetTyrLeuPheAsnHisGluIle 221
QY 508 -----ATTATGATTAAGTTATGAAAAACCAATACATCAATTCGACACTT 551
Db 221 eTyrLysLeuPheLeuPheHisAsnLysTyrLeuAsnAsnSerAsnIle---Proph 240
QY 552 AGGTAGAGAAATTACTTGATGCT-----GGTATTATTAACAGAAATCACACT 596
Db 240 eAsnLysAsnLeuIleGlnGluMetGluPheAsnLeuTyrTyrPheArgGlu-----Ile 258
QY 597 TAAACAGATTTTAATTATACGATTTTGTAAAGATTAATGATGATGATGATGAGC 656
Db 258 eLysAsnGluLysAsnTyrIleIle-----LysMetAsnLysLysGluIle 273
QY 657 CTATGACATGCTGTGAATGTTTGCAAACCTCACACCTGACACACTTACATTCATCA 716
Db 273 eTyrLys-----LysCySAsnValLysPhe---HisGluAsnValAspHisIleAs 289
QY 717 TAATGACGATGATATA----- 732
Db 289 PAsnGluLysIleLeuAsnIleLeuArgLeuTyrValAspAsnSerIleLeuAspIleAs 309
QY 733 -----TTAGTATGCTGCCATTCATTATAGTATATTTTCCAAATTTTGACTA 782
Db 309 PLeAsnAsnLysMetLeuCySAsnLysAsnAsnAsnLeuIleAsnGluAsnIleGluTyr 329
QY 783 T---AACAAATTAAACA-----TTTTCATGAATATTTATGAAATCTTACTGATTAATGA 833
Db 329 rIleSerLysLeuLeuAsnPhetYrCySThrLeuIleLysLysGluLysTyrAspAsnAs 349
QY 834 AATGACACGTTTTCAGTTACTTCACACCAATATACAGATTTTAAATATTTTACACATTA 893
Db 349 PmetThrIleTyrLysLeuLysGlu-----ValIleLysAlaIleThrHisH1 364
QY 894 TCATTTTCATGAT-----ATGAATTTTATGACTATATTAACATCTC---TATCGGG 944
Db 364 sIleLeuCySAspLysThrLysAsnLeuGluThrPheCySerAspIleAspTyrSerTh 384
QY 945 TGGTTTAAATATGTTAATACCAATATACATA---AACAACTAATGATGAGCC---TT 997
Db 384 rLeuLeuAsnSerLeuAsnAsnLysPheIleLeuAsnLysIleIleAspLysAsnPhel 404
QY 998 GTTTTTCATGACATTCATCGAGTA-----TC 1027
Db 404 eLeuPheTyrGluCySLeuLeuLysIleLeuLeuAsnIleLysPheValAsnPhelGln 424
QY 1028 CTTATGTGAT-----GTATCATGAAAAAATTCACACATGTTTACTTTTACGAACACT 1081
Db 424 rLeuCySLeSerLeuIleSerLeuLysAsnIleTyrTyrAsnIleLeuArgAsnAsnVal 444
QY 1082 ATTGCAACCAAGTTATCCCTACTTTTATGATGATGACAAATATTTTCATTATAT- 1140
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

Db 444 lTyr-----IleValAsnAsnValLeuPhe-AsnAspIleMetLysPheSerLeuTyrL 462
QY 1141 -----AAGATTGATAAAGTATTTTAAAGATGATTTTATA 1177
Db 462 eCySAsnIlePheLeuGluLysArgIleLysThrGluAsnGluAsnAlaValIleIle 482
QY 1178 TT-----AAATTAATCACGTGATATAC 1201
Db 482 lHisAsnAsnAspGlnThrAsnTyrSerAsnLysGluAsnIleLysAspIleIleLe 502
QY 1202 GTCAAATGATTGAAAAATAC-----TATTAATAT----- 1230
Db 502 lInLysArgIleLysGluTyrIlePheTyrLysMetGluAsnTyrLysAspPheHisPhe 522
QY 1231 -----GATATGATTAGCTTAATTCAT-----ACAATACATTAGAATGATTC 1276
Db 522 ySLeuLysAspSerAspLeuLeuSerIleLysLeuLeuSerAsnThrPheValLysIleA 542
QY 1277 AAGACAT-----ACGGTATGATTCGATC--- 1302
Db 542 sngluValTyrAsnSerTyrAspPheTyrLeuLeuPheAsnAsnIleSerCySleLeu 562
QY 1303 -----CATATACGTTAATTCGTTGTTTATATGATGATGATTCAT---G 1351
Db 562 yAsnPheLeuValAsnArgAsnSerValLysLysTyrLysAspThrTyrIleTyrIle 582
QY 1352 CACGTGATTTATTTTTCAAACACTATTTTAAACACAGCAAGTAAACAA- 1410
Db 582 eAsnAspLeuSerPheValTyrLysTyrIleLysAsnAsnAspArgThrLysLysLys 602
QY 1411 -----ATCATA 1417
Db 602 yAsnPhePheLeuLeuSerSerSerMetLysGluLeuIleCySAsnIleLeuSerV 622
QY 1418 TGACATGACCTTACGACTATACACTTACGATGATATACAGCAACACCATTCGAAATG 1477
Db 622 aIleSerAsnArgTyrIleLysHisLeuHisGluGluAspAsnPheAsp-----GlnLysA 640
QY 1478 AGGAGTTATGTTATCTAAAGCGTTTAAAGATTAATGACATACCTGATCGATCGTT 1537
Db 640 sPcGlnTyrValCySerLeuThrPheLeuAsnLeuPheAspLysIleIleHisP 660
QY 1538 CACATTTTAACTATATCCGTTAGATATACATGACATACATCAATC-----ATTAACG 1594
Db 660 heHisTyr-----IleTyrAsnLeuIlePcHisIsv 670
QY 1595 GTTACAAAACACTGACGTAAATATATATATTCCTACATTTGTCATCAGCTTCATGTT 1654
Db 670 aLlTyrLysThr-----TyrAsnTyrPheLysCySAsnLysLeuIleAsnGluAspIle 688
QY 1655 ATAACTATTTGTT-----CCTTCCATATCTTAACGGAAGGAAATTCAGC 1702
Db 688 lSerLeuLeuLeuLeuThrCySerLysPheGlnTyrPheIleGluAsnAsnSerAsnA 708
QY 1703 ACAATTTTATTTATTCGATACGTAGTATGTTATAGAA-----TCCG 1747
Db 708 sPArg-----TyrCySArgLysGluLeuLeuHisLeuLysTyrAsnIleIleAspAsp 726
QY 1748 TTGTTAAACCTTATTTAACCACCGCTTATTCGACCGGATAGCCTTAGTAATGGGATA 1807
Db 726 eulIleLysAsnTyrLeuAsn-----ThrTyrLysSerIleSerIle----- 739
QY 1808 TTGAAAACGACATGATAGATGTTTGA---CTGAATCATTAACAAATATGATCATAG 1864
Db 740 -----AspAsnIleSerLysIlePheIleSerLysSerAsnSerLysTyrThrCySg 757
QY 1865 AAGTGAATGGAAGATTAATAATGCTTCTGCTGATATACGGAATAACGCTTGCATACAA 1924
Db 757 lValAsnGluAsnLeuLeuLeuLysSer-----Leug 768
QY 1925 GCGTCGATTTTGAACCTTTGACGTGA-----CAATTCCTTTGAGC 1966
Db 768 lnsrGluPheGluLysValThrLysThrSerLysLysGlyGlyIleHisMetMetAspA 788


```

Db 1376 LeuValleuLeuIleAsnIleAlaSerPheTyr-----TyrThrLeuHisAsn 1392
QY ACCAAATCATATAACAAACACTATGATGAGCCTGTTTTCATATGACATCAATTCAGT 1023
Db 1393 AsnThrTyrThrSerSerTyrIleIleArg-----LysLysAspThrGlnArgGlu 1409
QY 1024 TATCCTATATGATGATATCATGAAAAAATTCACACATGTTATCTTTACGAACACTAT 1083
Db 1410 TyrGluLysIleIleLysGluLysLysIle-----IleGluHisAsn 1423
QY 1084 TCACAAACCAACG-----TTAAACCTCTACTTT-----TTAGATGAT 1119
Db 1424 AsnGlnLysAsnLysGluLysLeuIleAsnHisTyrGluAspIleAsnIleLeuAspGlu 1443
QY 1120 GACATATTAT-----T----- 1128
Db 1444 GluAsnPhelLysGlyAspHisLysAspIleLysValLeuLysLysTyrLysAsnGlyTyr 1463
QY 1129 -----TTTTCATATATAGATTCATTAAGATGATTTTAAACGATGATTTA 1173
Db 1464 TyrTyrSerLysIlePheSerLeuTyrProLeuAspGlnIleHisLeuAsn----- 1480
QY 1174 TTAATTAATTAATTAATCACTGATTCAGTCAATGATTTGAATATCTATATATATGAT 1233
Db 1481 ---IleGluLeuLysLys-----GluGluMetValAlaLysAspLysThrAsnGln 1496
QY 1234 AATGATTACGTTATATCATCAATAACATATACATTAAGAAATGATTCAGACATTCACGGTATT 1293
Db 1497 Gly-----AsnIleGlySerAsnLeuLeu-----LeuThrGlyAla 1508
QY 1294 GATTCATGATCATATACGTTAATTCGTTGTTATATATGAATGATTAATCTTCATGCA 1353
Db 1509 -----Ser 1509
QY 1354 CGTGATATATATTTTCAAACTATTTATTTAAACAGAGTAAGTTAA--AAC 1407
Db 1510 LysAspIleThrSerTyrLysAsnTyrTyrIleAspThrTyrIleLysMetGluLeuLeuLys 1529
QY 1408 AAATTCATATGACATCACTTACGACTATACATTCATGATGATATCAACAGAACACCCA 1467
Db 1530 LysLeuAsnIleLeuLeuProThrLeuTyr---IleLysGluIleLysAsnLysSerPro 1548
QY 1468 TACCA----- 1473
Db 1549 HisGluIleLysLeuSerSerMetAsnIleIleAspIlePheValSerLeuLysAsnVal 1568
QY 1474 -----AATGAGGAGGTTATG-----TTACTAAAGTCGTTTAAATGATTTAT 1518
Db 1569 LysIleArgAsnGluAspIleMetTyrLysLeuSerGlnLysTyrIleMetAspIlePhe 1588
QY 1519 GGCATACCTGATACGTTACATTTTACTTATTCGTTAGATGAATGAACATGAACATA 1578
Db 1589 -----PheHisAsnAsnLysValLysLeuGlnTyrGlnIleLysPhe 1602
QY 1579 TACAATATCATTAACGGTTACAAAAC--ACTGACGTAATATATATCTCTACATTT 1635
Db 1603 LeuAsnSerLeuThrPheLeuAspTyrIleLysGluAlaAspLeuPheLysThrPhe 1622
QY 1636 GTCACATCAGCTTCATG----- 1653
Db 1623 PhePheLysLysAsnLysIleAsnLysIleGlnLysGluLysLysGlnAsnAsn 1642
QY 1654 TATACCTATATGCTCTTCCTTCCATTAATAACGAAAGTGAAGATGACGACATTTAT 1713
Db 1643 TyrAsnLeuLeu-----TyrThrHisPheLeuLysIleProIle--HisAsnCysIle 1659
QY 1714 TATTCGCACTAGTACTGTTTATATGAATCCGTTGTAACCCCTTATTAACCCAGT 1773
Db 1660 TyrIleProAsnIleSerSerTyrIleLeuAsnPhelle-----Ser 1673
QY 1774 TTATTCGACCGGATAGCCTTAGTAATGGATATTTGAAACGAACAGATGATGAAGT 1833

```

```

Db 1674 IleTyrAspTyrPhe-----GluLysLysAspGlnTyr 1684
QY 1834 TTGTGACTGATCATATAGAAATATGATATGAGTGAATGAAAGATTAATAATTCCTCT 1893
Db 1685 ValIle-----TyrLysLysLeuLeuTyr----- 1692
QY 1894 GCTGGTATACGAAACAGCCCTTTGATACAGCGTCGATTTGAAACCTTTGTACGTGA 1953
Db 1692 ----- 1692
QY 1954 CAATCTTTGACGCTGCATTTATTTGAACATATAAGTATCTATATGCAAGTAC 2013
Db 1693 ---PheLeuAspGlnLysTyrLeuLysSerHisAsn-LysIleAsnSerMetAsnSerLeuAsn 1711
QY 2014 ATATCGA---TATATCGCTCTAAACTGAAATTTGATGTTGTTATGATATGATGAATAT 2070
Db 1711 pLysArgAsnIleIleLeuIleIleIleLeuLeuTyrIleSerSerProLeuAsnIle 1731
QY 2071 TTATACGATGACTTTATATGAACGCGAATTTATATTAACACCGTACAGAAATTC 2130
Db 1731 eleuSerIleArgLeu----- 1736
QY 2131 GACCAATGATCAATTCATGATTCCTTTATATGAAAGTGCATCGTTCATTT----- 2185
Db 1737 -----GlnThrLeuArgIlePheTyrTyrIleIleGlnSerAsnTyrPheSerLys 1754
QY 2186 -CACTTACGACCTTAT 2200
Db 1754 sHisAsnIleThrTyr 1759

```

RESULT 7
 AAB18288
 ID AAB18288 standard; Protein; 1182 AA.
 AC AAB18288;
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:146.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFF/) HOFFMAN S.
 PA (CARD/) CARDCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 DR WPI; 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 XX
 PS Disclosure; Page 347-350; 577pp; English.
 CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)
 CC vaccines against P. falciparum infection comprising (i) or (ii).
 CC (i) and (ii) are useful for the development of vaccines against


```

Db 617 GluLeuLeuPheLeuLeuLeuAsnTyrArg-----ProSerLeuLysGlnArgAsn 634
      ::::::::::::::::::::|
QY 1543 -----TTTAACTTATTCGCTTAGATGAATAACAATGACTATAC-- 1581
      |||
Db 635 LysArgAsnAsnLeuSerLeuLeuAsnAsnLeuTyrLeuLysAsnLeuLysLysTyrIle 654
      |||
QY 1581 ----- 1581
Db 655 LysLysLysLysLysLysLysLysTyrIleTyrIleTyrIleCysLysLysLysAsn 674
      ::::::::::::::::::::|
QY 1582 AATATCATTAACGGTTACAA-----AACACTGAACCTAATATATTA 1623
      |||
Db 675 AsnValGlyAsnLeuHisLysHisAsnValMetMetThrSerAsnHisAsnAsnLeuLeu 694
      |||
QY 1624 TTCCTCATTTGTCACATACGTCATCTGATTAACCTATTTGGTCTCTCCAACTTA 1683
      |||
Db 695 PheArgSerPheGluTyrValLys--ValHisLysLeuLeuLeuPheLeuAsnLeuLeu 713
      |||
QY 1684 ACGGAAGTGAATGACGACAAATTTATTTATTCGATCTGATAGTTGATATGAAA 1743
      ::|
Db 714 IleLysSerAsnLeuTyrIleAsnTyrGluTrp-----SerLeuTyrPheLeu 729
      |||
QY 1744 TCCGCTGTATA-----CCCTTATGACCCCGCTTATTCGACCCGATAGCCTTA 1794
      |||
Db 730 SerLeuLysGlnLysHisAlaPheLeuLysLysGlyPheTyrIleLeuCysTyr 749
      |||
QY 1795 GGTAAATGGGATTTGAAACGAAACAGATA-----GATAGAGATTTGTA 1839
      ::|
Db 750 IleLeuPheHisIleGlnAsnAsnHisIleLeuTyrLysSerTyrGluHisIlePheAsn 769
      |||
QY 1840 CTGAATCATAGAAATATGATATGATGAAGTGAATGAAATGAAATTA----- 1884
      ::|
Db 770 ProTyrAsnLysTyrAsnLeuTyrAsnLeuLysAsnLeuLeuLysCysThrLeuProGln 789
      |||
QY 1885 -----ATTGCTCTGCTGCTATACGAAA 1908
      |||
Db 790 IleLeuGlyThrSerAsnLeuTyrSerLeuLeuTyrValAlaPheLeuTyrSerThrAsn 809
      |||
QY 1909 AACCCCTTGTATACACGCTGATTTGAAACCTTTGTACGACACATCTTTGACGGT 1968
      |||
Db 810 AsnThrIleAsnPheLeuLysIlePhePheThrIleLeu--GlnLysPheTyrAspSer 828
      |||
QY 1969 GCCATT-----ATTGAAACATTAAGTATCTATATATGACGACAGTACATATCG 2019
      ::|
Db 829 SerMetIleLysGlnIleGlnAsnAspLysAsnAsnTyrGlnHisIleSerCysHisAsn 848
      |||
QY 2020 ATATATCCGCTCTAAACTGAATGTATGTTGATGATGATGATGATATTTTACTGAT 2079
      |||
Db 849 TyrSerProLysLysAspAsn-----SerGluTyrTyrIlePro 861
      |||
QY 2080 GAACCTAATATGAACGTAATTTATATTAAGACGCTAGAGAAAATTCGACCATAGT 2139
      ::|
Db 862 AspAspHisAsnLysLeuLeuTyr-----AsnTyrSerTyrAsn 874
      |||
QY 2140 CAATTGTGATTTATCTTTATTTATGAAGTACATCGTTCATTTTCACTAAGACTTA 2199
      |||
Db 875 GlnLeuTyrGlnLysAsnHisPheAsnAspAsp-----AsnIlePheLeuHisAspLeu 892
      |||
QY 2200 TTTCAGTTGAAGCTGCTGATCATATACAAA 2229
      |||
Db 893 LysIleTyrGluArgAsnIleAsnAsnLys 902
      |||

```

```

RESULT 8
AAB16580
ID AAB16580 standard; Protein; 42 AA.
AC AAB16580;
XX
XX 27-OCT-2000 (first entry)
XX
XX Bacteriophage 44AHD protein sequence 44HDOFR047.
XX

```

```

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; bacterial infection.
OS Bacteriophage 44AHD.
PN WO200032825-A2.
XX
XX 08-JUN-2000.
XX
XX 03-DEC-1999; 99WO-1B02040.
XX
XX 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
DR WPI; 2000-412361/35.
DR N-PSDB; AAB69065.
XX
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX
PS Example 9; Page 279-280; 456pp; English.
XX
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAB68243 to AAB69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SQ Sequence 42 AA:
Alignment Scores:
Pred. No.: 2,296-12 Length: 42
Score: 225.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 5,67% Indels: 0
DB: 21 Gaps: 0
US-09-727-892a-2 (1-2286) x AAB16580 (1-42)
QY 717 ATGAATGTATGTAAGTTGTCAGGTGAGATTGCAAAACATTTGACAGCATAGTCANA 658
      |||
Db 1 MetAsnValCysLysLeuPheArgCysGlnPheCysLysThrPheHisSerIleValIle 20
      |||
QY 657 GGCTTCACATTCATCATATCATATTCATTTATCAAAATGCTATTAATTAATCTGTTT 598
      |||
Db 21 GlyPheThrIleLeuHisIleIleIleIlePheLeuLysAsnArgIleLeuLysIleCysPhe 40
      |||
QY 597 AAGTG 592
      |||
Db 41 LysLeu 42
      |||

```

```

RESULT 9
AAB16603
ID AAB16603 standard; Protein; 573 AA.
AC AAB16603;
XX
XX 27-OCT-2000 (first entry)
XX
XX

```

XX Bacteriophage 192 protein sequence 182ORF002.
 DE Bacteriophage: antimicrobial; genome; identification; antibacterial.
 XX Bacteriophage: growth inhibition; bacterial infection.
 KW Bacteriophage 182.
 OS Bacteriophage 182.
 XX WO200032825-A2.
 PN 08-JUN-2000.
 XX 03-DEC-1999; 99WO-1B02040.
 PF 03-DEC-1998; 98US-0110992.
 PR 03-JUN-1999; 99US-0326144.
 PR 28-SEP-1999; 99US-0407804.
 PR 30-SEP-1999; 99US-0157218.
 PR 01-DEC-1999; 99US-0168777.
 PR 02-DEC-1999; 99US-0454252.
 XX (PHAG-) PHAGE TECH INC.
 PA Pelletier J, Gros P, Dubow M;
 PI WPI, 2000-412361/35.
 DR N-PSDB; AAA69089.
 XX
 PT Identifying a bacteriophage coding region for treating bacterial
 PT infections comprises identifying a nucleic acid encoding a product that
 PT inhibits bacteria when a bacteriophage infects a bacterium
 XX
 PS Example 13; Page 318-319; 456pp; English.
 CC The present invention describes a method for identifying a bacteriophage
 CC coding region encoding a product active on an essential bacterial
 CC target. The method comprises identifying a nucleic acid sequence encoding
 CC a gene product that provides a bacteria-inhibiting function when an
 CC uncharacterised bacteriophage infects a pathogenic bacterium. The
 CC compound active on a target of a bacteriophage inhibitor protein in a
 CC bacteria is used to treat or prevent a bacterial infection in an animal.
 CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
 CC nucleotide and protein sequences which are used in the exemplification of
 CC the present invention.
 XX
 SO Sequence 573 AA:
 Alignment Scores:
 Pred. No.: 1,39e-11 Length: 573
 Score: 220.00 Matches: 128
 Percent Similarity: 34.43% Conservative: 91
 Best Local Similarity: 20.13% Mismatches: 229
 Query Match: 5,41% Indels: 188
 DB: 21 Gaps: 30
 US-09-727-892a-2 (1-2286) x AAB16603 (1-573)
 QY 205 GAGCATTTTATCGATGTGAAAGCGTATACATCAAAATCAAAACAGATATT 264
 DB 41 AspSerPhehegIutrypyslys-----MetGlnIysertHraspile 55
 QY 265 ATCATGATTCGACATACGTATATAATACATATCATTTTACTT----- 312
 DB 56 TyrPhe-----Hisasn---GluysPheaspGlygluPheMetLeuSerTrpLeuphe 72
 QY 313 AAAGACACATGCGTATTTGATATATATACACGCGAAATATATAT----- 360
 DB 73 LysAsnGlyPheLysTrpCysLysGluAlaLysGluAspArgThrPheSerThrLeuile 92
 QY 361 -----TTAAATCTGCAGAGAAATGACACACATTAATA 396
 DB 93 SerAsnMetGlyGlnTrpTyrAlaLeuGluIleCysTrpGluValAsnTyrThrThrThr 112

QY 397 ATGAAGAGGCTACTATTATTAGCCAAAATCAAAATGATTTAGAAAACGTGTAA 456
 DB 123 LysSerGlyLysThrLysLysGluLysSerArgThrIleLeuIleTyrAspSer----- 129
 QY 457 TCTTCAATCATTTAGATTTAACAATGTTTAAATGGTTTAAATTTAATATTATGAT 516
 DB 130 -----LeuysLysTyrProPheProVal----- 137
 QY 517 AACTTTATGAAAACCATACATTCATTCACATATAGTAGAATATGCTGATGTGT 576
 DB 137 ----- 137
 QY 577 TATTTACAGATACCACTTAAACAGATTTTATATATAGATTTTGTATTAAGATAT 636
 DB 138 -----LysGlnIleAlaGluAlaPheAsnProIleLysLysGlyGluile 153
 QY 637 GATATGATGATAGAGAGCCATGACATACATGCTGTGAAATGTTTGCAAAATCACACT 696
 DB 154 AspTyrThrLysGluArgProIleGlyTyr-----LysProThrLys 167
 QY 697 GAACACTTACATACATTCATATATGACGTATATATATGATGTCATATTCATTTAT 756
 DB 168 AspGluTrpGluTyrLeuLysAsnAspIleGlnIleMetAlaMetAla----- 183
 QY 757 ACTGATATATTCCAAATTTGACTATACAAATTAACATTTTATGATGATATTATGGA 816
 DB 184 -----LeuLysIleGlnPheAspGlnGlyLeu----- 192
 QY 817 TCTTACTGAAATGAATGACACGT---TTTACGTATACCAACCAATACAGATATT 873
 DB 193 -----ThrArgMetThrArgLysSerAspAlaLeuGlyAspTyrLysAsp--- 207
 QY 874 AAATATCTTATACACAT-----TATCATTTCCATGATATG 909
 DB 208 TrpLeuLysAlaThrHisGlyLysSerThrPheLysGlnTrpPheProIleLeuSerLeu 227
 QY 910 AATTTTTATGACTATATTAATTCATTTTCATGCGTGTGTTAAATGATATACACCAA 969
 DB 228 GlyPheAspLysAspLeuArgLysAlaTyrLysGlyGlyPheThrTrpValAsnLysVal 247
 QY 970 TACATTAACAACTAATGATGAGACCTGTTTCTATGACATCAATTCGAGTTATCT 1029
 DB 248 PheGlnGlyLysGluIleGlyAspGly---IleValPheAspValAsnSerLeuTyrPro 266
 QY 1030 TATGTGATGATTCATGAAAATTCGACATGTTA---TACTTTGAGAACACTATATCA 1086
 DB 267 SerGlnMetTyrValArgProLeuProTyrGlyThrProLeuPheTyrGlyGlyTyr 286
 QY 1087 GAACCAACGTTAATCCCTACTTTTATGATGATGACATTTATTTTATATATAGATT 1146
 DB 287 LysPro-----AsnAsnAspTyrProLeuTyr----- 295
 QY 1147 GATAAGATGATTTTACGATGATTTATTAATTAATTAATGATGATGATGATGATG 1206
 DB 296 -----IleGlnAsnIleLysValArgPheArgLeuLysGlyTyrIleProThr 312
 QY 1207 ATGATTTGTAATATCTATAAT-----AATGATATGATTTTCTTATATACATACAAT 1260
 DB 313 IleGlnValLysGlnSerSerLeuPheIleGlnAsnGluTyrLeuGlnLysSerValAsn 332
 QY 1261 ACATTAAGATGATTCAA-----GACTATACGGGTATGATTCGATGATGATGAT 1308
 DB 333 LysLeuGlyValAspGluLeuIleAspLeuThrLeuThrAsnValAsp---LeuGlnLeu 351
 QY 1309 CGGTATATTCGTTGTATATGATGATGATGATGATGATGATGATGATGATGATG 1353
 DB 352 PhePheGlnHisTyrAspIleLeuGlnIleHisTyrThrTyrGlyTyrMetPheLysAla 371
 QY 1354 CGTGATATTTATTTTCAAAATCATTTTATTA-----ACACAA 1392
 DB 372 SerCysAspMetPheLysGlyTyrIleAspLysTrpIleGluValLysAsnThrThrGlu 391
 QY 1393 GGTAGTTTAAAAAACAAATATCATATGACATACCTTACGATCATCATTAAGATGAT 1452


```
Db      846 LysAsnIleTyr-----AsnAsnAspSPleLeysLys 856
      ::|||
      ::|||
RESULT 11
ID      AAB16585 standard; Protein: 40 AA.
XX
XX      AAB16585;
XX
XX      27-OCT-2000 (first entry)
XX
XX      Bacteriophage 44AHD protein sequence 44HDDRF051.
XX
XX      Bacteriophage; antimicrobial; genome; identification; antibacterial;
XX      bacterial growth inhibition; bacterial infection.
XX
XX      Bacteriophage 44AHD.
XX
XX      WO200032825-A2.
XX
XX      08-JUN-2000.
XX
XX      03-DEC-1999; 99WC-IB02040.
XX
XX      03-DEC-1998; 98US-0110992.
XX
XX      03-JUN-1999; 99US-0326144.
XX
XX      28-SEP-1999; 99US-0407804.
XX
XX      30-SEP-1999; 99US-0157218.
XX
XX      01-DEC-1999; 99US-0168777.
XX
XX      02-DEC-1999; 99US-0454252.
XX
XX      (PHAG-) PHAGETECH INC.
XX
XX      Pelletier J, Gros P, Dubow M;
XX
XX      WPI: 2000-412361/35.
XX
XX      N-PSDB; AAB69070.
XX
XX
XX      Identifying a bacteriophage coding region for treating bacterial
XX      infections comprises identifying a nucleic acid encoding a product that
XX      inhibits bacteria when a bacteriophage infects a bacterium
XX
XX      Example 9; Page 280; 456pp; English.
XX
XX      The present invention describes a method for identifying a bacteriophage
XX      coding region encoding a product active on an essential bacterial
XX      target. The method comprises identifying a nucleic acid sequence encoding
XX      a gene product that provides a bacteria-inhibiting function when an
XX      uncharacterised bacteriophage infects a pathogenic bacterium. The
XX      compound active on a target of a bacteriophage inhibitor protein in a
XX      bacteria is used to treat or prevent a bacterial infection in an animal.
XX      CC AAB6843 to AAB6942 and AAB16523 to AAB16954 represent bacteriophage
XX      nucleotide and protein sequences which are used in the exemplification of
XX      the present invention.
XX
XX      Sequence 40 AA;
XX
XX      Alignment Scores:
XX      Pred. No.: 9 31e-12 length: 40
XX      Score: 218.00 Matches: 40
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 5.36% Indels: 0
XX      Gaps: 0
XX
US-09-727-892a-2 (1-2286) x AAB16585 (1-40)
OY      1562 ATGATACATGACCTATACATATCATTAACGCTTCAAAAACACGACGATATATAT 1621
Db      1 MetIlethrmelAsnTyrThrIleSerleuThrValThrIleuAsnValIleTyr 20
OY      1622 TATCTCTACATTTGTGCACACGCTTCATGTAATTAAGTTTGGTTCCTTCCAACT 1681
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      21 TyrSerLeuHisIleuSerHisValHisCysIleThrTyrTrpHeuSerAsnTr 40
      ::|||
      ::|||
RESULT 12
ID      ABB77616 standard; Protein: 767 AA.
XX
XX      ABB77616;
XX
XX      01-JUL-2002 (first entry)
XX
XX      AMEPV early transcription factor-large protein.
XX
XX      AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
XX      genetic deficiency disorder; transcription factor large-protein.
XX
XX      Amsacta moorei entomopoxvirus.
XX
XX      WO200212526-A2.
XX
XX      14-FEB-2002.
XX
XX      10-AUG-2001; 2001WO-US25287.
XX
XX      10-AUG-2000; 2000US-224479P.
XX
XX      14-SEP-2000; 2000US-0662254.
XX
XX      (UYFL ) UNIV FLORIDA.
XX
XX      Moyer RW, Li Y, Bawden AL;
XX
XX      WPI: 2002-227161/28.
XX
XX      N-PSDB; ABL56219.
XX
XX      Novel recombinant entomopox virus vector useful for delivering
XX      polynucleotide encoding protein to vertebrate cell, comprises
XX      polynucleotide encoding protein operably linked with heterologous
XX      promoter sequence
XX
XX      Claim 75; Page 270-273; 326pp; English.
XX
XX      The invention relates to a recombinant entomopox virus (EPV) vector,
XX      comprising a polynucleotide encoding a protein operably linked with a
XX      heterologous promoter sequence. The invention also concerns methods for
XX      providing gene therapy for genetic deficiency disorders. Vectors of the
XX      invention are useful for delivering a polynucleotide encoding a protein
XX      to a vertebrate cell preferably a mammalian cell, such as a human cell.
XX      The vector is introduced into the vertebrate cell by infection in a viral
XX      particle, or by transfection, transduction, or infection either in vitro
XX      or in vivo. The vector is useful for the delivery and expression of
XX      biologically useful proteins in gene therapy protocols, and for
XX      delivering large DNA segments for engineering of vertebrate cells.
XX      CC polynucleotides of the invention have applications in techniques such as
XX      their use as insertion sites for foreign genes of interest, hybridisation
XX      probes, for chromosome and gene mapping, in PCR technologies, and in the
XX      production of sense or antisense nucleic acids. Vectors of the invention
XX      provide for stable integration and expression of heterologous DNA in host
XX      cells, and are adapted for accepting large heterologous polynucleotide
XX      inserts which can be delivered in an infected or transformed cell and
XX      expressed in a stable fraction. The current sequence represents an
XX      amsacta moorei entomopoxvirus (AMEPV) early transcription factor-large
XX      protein (AMV105).
XX
XX      Sequence 767 AA;
XX
XX      Alignment Scores:
XX      Pred. No.: 2 31e-11 length: 767
XX      Score: 218.00 Matches: 168
XX      Percent Similarity: 36.17% Conservative: 117
XX      Best Local Similarity: 21.32% Mismatches: 272
XX      Query Match: 5.36% Indels: 232
XX      Gaps: 37
XX
US-09-727-892a-2 (1-2286) x ABB77616 (1-767)
```

179 TATTTCGAGTTTCGATCTTTTATGAGCCATTTATAGTATGCGAAAGACGTGATA 238
104 TTTTTSnAsnSngluillepAspProTyrAsnillelysAsn---Asnleuile 122
239 CAATTCACAAATCAAAACAGATATTATCATGATGCACATACCTAATTAATACGATA 298
123 GluSerSnspsnlyslsyleleuYrmetLeuAspIleSerIlethrSnglyala 142
299 ATCATTTTACTTTAAAGACACATGCGTTATTATTATTAATTTTACCGCGAAATATAT 358
143 Ile--PheCysValThrThrAsnSerTyrThrAsnThrAsnleuAlaLysGluGlyIleT 162
359 ATTAAATTCGAGAGAAATGACACACATTAATAATTAAGAGCGGTACATTTTNG 418
162 YrSerlySile-----TyrThrGluTyrIleGluGluIleIlePheAsnI 177
419 CCAAAATCAAAATGTAATTTTAGAAAAAGTGTAAATCTTCATCAATTTAGATTAA 478
177 LetTyrSAsnAsnTyrLysLeuSerSerValValLysGluSerGluGluTyrSerLeuT 197
479 CAAATGTTTAAATGGTTT-----AAATTTATATTATGATTAATCTTATATA 526
197 hr-----AsnAsnPheAspPilellelysLeuSerAsnIleAsnLysTyrLysL 214
527 AAACCAATCATCATTTGCAACATTTAGTAGAATTAATTAATCTGATTTATTAACAG 586
214 Ystrhr---leucylleGlyValTyrAspLysTyrTyrIleLysGlyAspLysIleSerI 233
587 AATCACAACCTAAACAGATTTTAATTAACG----- 618
233 leleuAspAsnTyrAsnAspSerGluTyrThrSerLeuTyrIleTyrIleAspGluAsnA 253
619 --ATTTTTGTATAAGATTAATGATATGATGATGATGACCATGATGCTGCGAAT 676
253 snllelleLysIleThrAsnAspValleuIlethrGluLysLeuThrTyrPheThrAspI 273
677 GTTTTCGAAA-----CTCACACCTGGAACAACCTTA 706
273 leleuLysGluGluGluIleLysAsnIlelleLysSerThrSerProLysSerIleI 293
707 CATACATCATATGACGATATTATATTAGTATGCGATATTCATTTATGATATAT 766
293 letTyrIleTyrPheAspPheThrPheLeuAspSerAsnIleAsnIleGluTyrAsp----- 310
767 TTCCAAATTTTGAATTAACAATTAACATTTTCAATTTGAATATTATGAAATCTTACTTGA 826
311 -----LeuLysPhePheLeuAsnVal-----T 318
827 AATATGAATGACAGCTTTTTCAGTTTCTCAACCAATATACAGATATTAAATATCTTATA 886
318 hrAsnThrArgAsnIlePheIleAspMetSerTyrLysIleAsnIleMetThrSerLysA 338
887 CACATATCATTTTCCATGATGATGATTTTATGACTATATTAATCATCTATCGGTG 946
338 snhIleSerPheArgSerPheAsn----- 346
947 GTTTAATATGTTATACACCAATATACATAACAATAATGATGACCTGTTTTCCTA 1006
347 --lleAspValAsnleuLysLysTyrLeuSerLeuIle-----L 360
1007 TTGACATCATTTGATGATTCCTTATGATGATGATGATGATGATGATGATGATGATGAT 1060
360 euclLysThrSnhIlePheAsnLysIleGluLysHisAlaArgLeuLysLysIleAspG 380
1061 GGTATATAC-----TTTACGAACACTATTCAGAACCAAGTTAATCCCTACTTTT 1111
380 lueuLysrProSerArgTyrCysGluAsnTyrLysAspValLysArgInProValLeuI 400
1112 TGAT-----GATGACATTAATTTTTCATTTATTAAGTTATTAAGTATTAAGT 1165
400 leAspSerIleAspGluAsnTyr----- 407

1166 ATGATTTATTAATTAATTAATACAGCTGATTTACGTCAATATGATTTGTAATATACATA 1225
408 -----LeuIleLysIleSerAsp-----LysTyrTyrV 417
1226 AATATGATTAAGATTA-----GTTATATATCAATACAAATATACATTAAGATTCACAGAC 1281
417 alGlyLysGluAspPheThrArgThrTyrGlnHisLysGlyThrLysLysIlePheAsp 437
1282 ATTACGGGTATTTGATTCATGATGATACATATACGTGTAATTCGTTGTTATATGAATGTCAA 1341
437 rotYrLysTyr-GlyAspValTyrIleAspAspAsnGly---LeuIleTyrGluCysSer 455
1342 TACTTTCATGACGCGATATATATTTCAAACATTTATTAATAACCAAGT----- 1395
456 SerIleTyrTyrSerAsnMetGlyPheLeuAsnAsnIleTyrLeuAlaSerGlyLys 475
1386 -----AACTTAATAAACAATAATCAATATGACATACCTTAC 1431
476 ThrCysTyrProCysCysTyrSerLysGlnLysAsnArg----- 488
1432 GACTTCACATTTACGATGATATACAGACACACCATACCAATGAGAGGTATATGTA 1481
489 -----AspGluIlePheGluSerCysValTyrAsnLysGluIleleu 503
1492 -----TCTAAGTCGTTTA-----AATGCA 1512
504 GluAspLysIleAsnProIlelleValAsnTyrGlyArgIlelleLeuSerLysAsnGly 523
1513 TTATATGCCATA---CTCGCATACGTTCA----- 1539
524 leuSerLysLeuSerProLysLeuAsnAsnIleleuAsnAlaAsnSerLysIleAspIle 543
1540 -----CATTTTAACTTTATTCGCTTATAGATGATGATCAATGATCAATGATCAATG 1593
544 VallyshIsthrAsnArgIleAspPheSerAspSn-----TyrThrIleIleMet 560
1594 GGTACAAAACACTGAAAGTAAATATATATTCATCATTTGTCACATCACTGTCATG 1653
561 SerTyrGlnProThr-----IleThrIleArgAsnPhe 571
1654 TATACATATTTGTTTCCTTCATTAATTAAGGAAAGTGA-----ATTGACAGC 1704
572 AspAspMet-----TyrTyrPheIlelleAsnAsnAlaIleValIleAsnAsp 588
1705 AATTTTATTTATTTGCGATACGATAGTTTGTATATGAATCC----- 1746
589 AsnIleValTyrThrAspLysSerIleleuLysMetAsnAsnAsnIleAsnValPhe 608
1747 -----CTTGTTAAACCTTATTTGAACCCAGTTTATTCGAC 1782
609 llellelleGluAsnArgIleHisGluLeuLysAsnIleAspLysGlnSerLysTyrAsp 628
1783 CCGATAGCCTTAGGTAAATGGGATATGGAACAGACAGATATAGATGTTTGTACTG 1842
629 AspIleValIleAsnLysIleAsp-----AspLysLysIleLysIle 642
1843 AATCATTAAGAATATGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1902
643 lleLysLysTyrPheAsnIlelleSerAsnIleleArgAsnProIleSerAsnAsnGlyIle 662
1903 CCGAAAAAGCGCTTGTATACAGCGTGCATTTTGAACCTTGTACGTGAAACAATCTTT 1962
663 -----SerIleThrAspAspAlaLysThr-----Ile 671
1963 GACGGTGCCATTTATGAAAC---AATPAAAGTATCTATATATGAGCAAGGTATACATTCG 2019
672 AspGlyGluIleuIleGluAsnLysAsnIleLysTyrPheSerGluTyrAsnAsnIleSer 691
2020 AATATACGCTATAACGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 2079
692 leuLysProLysSerThrSer----- 698
2080 GAACCTAATATGAACAGCTGATTTATATTAATAAGACCTAGAGAAATTTGACCATAGT 2139

QY 838 ACA-----CGTTTCAGTACTCTACACCAATATCAAGATTATAAANA 879
 Db 1018 ValGlnAsnTyrLysAsnGlnLysAsnThrAsnMetGlnHisTyrAsnGlnLysLeu 1037
 QY 880 ---TCCTATACACATTTCATTCATCCATGATGATGATTTATGACATATTAATCAATTC 936
 Db 1038 PheLeuTyrProLeuTyrTyrLeuGlnLysAsnTyrPhe----- 1051
 QY 937 TATCGTGGTGGTTAAATATGATTAACACCAATACACCAACTAATGTAGCGCT 996
 Db 1052 -----LeuAsnValValAsnHisLeuPheAsnLys----- 1062
 QY 997 TGTTCCTATGACATTCATCGAGTATCCCTATGTGATGATATGACAAATAATTC 1056
 Db 1063 -----AsnTyrAsnAsnThrPhePheTyrThrCysGlnLeuSer 1078
 QY 1057 ACATGGTATACCTTTACGACACTATTCGACCAACGTTAATCCCTACTTTTACAT 1116
 Db 1079 LysGlyLeuTyrTyrPheLeuAsnTyrTyr-----ThrLeuLeu 1092
 QY 1117 GATGACATATTTTCATATATATATGATTAAGATGTATTAACGATGATTTATTA 1176
 Db 1093 SerSerAsnTyrLysAla-----GlnGlnLeuLysThrAspAsnLys 1107
 QY 1177 ATTAATTT----- 1185
 Db 1108 CysAsnLysAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 1127
 QY 1186 ---AAATCAGCTATATACCTCAATGATTTGTAATTAATATATATATGATTTAC 1242
 Db 1128 AsnAsnAsnAsnAsnAsnAsnTyrAsnAsnAsnAsnAsnValTyr 1147
 QY 1243 GTAATATCAATCAATCAATCAATTAAGATGATTCACAGACATACG-----CGT 1290
 Db 1148 ProLeuLeuAsnHisPheThrThrPheTyrGlnMetValThrTyrLeuLysAsn 1167
 QY 1291 ATTGATTCATGCAATTA----- 1308
 Db 1168 IleTyrArgGlnHisIleSerLysPhePheTyrIlePheValAlaLeuSerLysPhePhe 1187
 QY 1309 -----CGGTGAATTCG----- 1320
 Db 1188 LeuMetAsnSerTyrGlnGlnSerAsnThrAsnLysArgGlnLysSerIleHisMetGln 1207
 QY 1321 -----TTGTATATATGATGATGATGATTTGATTCATGACGATGATAT 1362
 Db 1208 AsnValLeuTyrIleLeuTyrIleIleArgLysGlnTyrGlnHisValLysSerIle 1227
 QY 1363 ATTTT-----CAAACTATTTTATTAAACACAGTAAGTTAAACAAA 1410
 Db 1228 LeuTyrAspLysSerAsnGlnLysAsnTyrPheArgPheAsnGlnLysAspIleLysMet 1247
 QY 1411 ATCAATATGACATCACCTACGACTATCACATT-----ACTGATGATATCAAC 1458
 Db 1248 GlnAsnThrAsnMetLeuTyrAsnIleIleLeuAsnAsnPheSerThrGlnLysPheLysAsp 1267
 QY 1459 GAACAC-----CCATCTCAAAATGAGAGGTTATGTTCTTAAGCTCGTTTAAT 1509
 Db 1268 GlnPheMetThrLeuGlnLysAsnAsnGlnLysAsp-----AsnAsnLysMetIleLeuAsp 1285
 QY 1510 GGATTTATGACATACCTGATACCTGATTCATGATTTAATCTTATCCGTTATGATGATAC 1569
 Db 1286 AsnIleAsnValAspAsnIleAsnAspLeuIleLysSerHisHisCysAspAsnAsn 1305
 QY 1570 -----AATGACTATACAAAT----- 1584
 Db 1306 LysLysGlnAspThrSerSerLeuHisAsnLysLeuTyrAsnGlnLysHisPheLeuIle 1325
 QY 1585 ---ATCATTAACGGTTTACAAAACACTGACGATATATATTTCTGACATTTGTACACA 1641
 Db 1326 MetPheLeuAsnAsnTyrLeuAspAsnThrLysHisPheLysIleAsnHisPheLeuSer 1345

QY 1642 TCACGTTCAATGTATTAACCTTATGGTCCCT-----TTCCAAATACCTTAACGGA 1689
 Db 1346 SerLeuPheTyrIleAsnLysIleIleProPheAsnMetLysHisMetTyrHisLeu 1365
 QY 1690 AGTGAATGACGACAAAT----- 1722
 Db 1366 ThrTyrLeuHisLysAsnHisLysIleTyrLysAsnLysPhePheTyrIleTyrAsnGly 1385
 QY 1723 ACTGAT-----AGTTGATATGAAATCCGCT 1749
 Db 1386 LeuAspLeuLeuLysSerTyrLeuValHisIleLysLysLeuTyrIleAsnSerTyr 1405
 QY 1750 GTTAACCCCTTTTGAACCCCGCTTATTCGACCCGATAGCCTTAGTAATGGAATAT 1809
 Db 1406 IleLysSerTyrAsnAsnLysLysLysAsnAsnVal-----AsnGlyAspVal 1422
 QY 1810 GAAAC-----GAACAGATAGATTAAGATGTTGTA--- 1839
 Db 1423 TyrAsnAsnPheMetTyrLysTyrAsnLysLeuTyrAspAsnIleAspTyrIlePheLeuLys 1442
 QY 1840 -----CTGAATCATTAAGAA-----TATGATAT 1863
 Db 1443 LysLysAsnLeuPheCysTyrTyrThrAsnHisLeuSerLeuLeuTyrPheThrTyrIleTyr 1462
 QY 1864 GAAGTGAATGGAAGATTAATAATTCCTCTGCTGCTATACCGAAACGCTTTGATACA 1923
 Db 1463 SerLeuAsn---LysPheTyrTyrCysThr---LeuTyrTyrAsnLysSerLysCysPheTyr 1481
 QY 1924 AGCGTCGATTT---TTGAAACCTTTGTACGTGAACAAATCTTTGACGCGTCATTATGAA 1960
 Db 1481 TTYrLysIleAsnIleGlnAsnIleHisPheLysAsn-----Ly 1494
 QY 1981 AACATTAAGATCATTAATGACGACGATACATATGATATACGCTCA----- 2032
 Db 1494 IleIleLeuPhePheIlePheThrGlnCysLysTyrIleTyrIleLysPhePheArgLe 1514
 QY 2033 -----AACTGAATTTGTA 2046
 Db 1514 LeuValGlnSerIlePheSerSerGlnPheGlnLysValGlyLysLeuSerLeuTyr 1534
 QY 2047 TGTGTAATGTATATGATGATATTTTACTGATGA-----CT 2084
 Db 1534 Ttle-----LeuSerAsnIleIleLeuLeuLeuValLysAsnSerArgMetLysLe 1551
 QY 2085 TAAATGAACGTCGAATTTATTTAAAGAC---GCTAGACAAATTTTCGACCATATGCA 2141
 Db 1551 AsnIleLysLysLysIleIleLysAsnIleSerLysHisIleTyrSerAsnAsnGln 1571
 QY 2142 ATTT-----GATGATATCTTTATATTTGAAGTACATCGCTTCATTTTACTTAA 2192
 Db 1571 upHeLysAsnAsnLysIleLysLysIleHisThrAsnAsnAsnSerMetSerLysAs 1591
 QY 2193 C-----GACTT 2198
 Db 1591 LeuPheIleCysAsnLysLeuAsnIleGlnTyrAsnTyrIlePheProMetAspLe 1611
 QY 2199 ATTTCACGTTTAACGTTGACGATTAACAAATCGATTTGCATATATTAACGTCGACA 2258
 Db 1611 upHe---LysSerSerLeuSerHisGlnThrGlnLeuIleIleAsnLysLeuGln 1630
 QY 2259 TGATGAATTAATAAAAGCAAC 2280
 Db 1630 AsnIleLeuAsnAsnAsnAsn 1637
 RESULT 14
 AAB16581
 ID AAB16581 standard; Protein: 42 AA.
 AC AAB16581;
 XX
 XX 27-OCT-2000 (first entry)
 DT
 DE Bacteriophage 44AHD protein sequence 44HDBRF045.

```
XX Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; bacterial infection.
XX
OS Bacteriophage 44AHJD.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX
DR WPI; 2000-412361/35.
DR N-PDB; AAA69066.
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX
PS Example 9; Page 280; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SQ Sequence 42 AA;
XX
Alignment Scores:
Pred. No.: 3 58e-11 length: 42
Score: 212.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.35% Indels: 0
DB: 21 Gaps: 0
US-09-727-892a-2 (1-2286) x AAB16581 (1-42)
QY 1973 ATGGCAGCGTCGAAGAATTGTCACGTACAAAGGTTCAAAATCGCGCTGTATCAAG 1914
Db 1 MetAlaProSerLysAsnCySerArgThrLysValSerLysSerThrLeuValSerLys 20
QY 1913 GCGTTTTCGGTATACAGCAGGAAGCAATTTTATCTTCATTCACCTCATATGCAAT 1854
Db 21 AlaIhepHeGlyIleProIaIaIaIleuIlePheProPheThrSerTyrAlaTyr 40
QY 1853 TTCTTA 1848
Db 41 PheIeu 42
RESULT 15
AAV31947
ID AAV31947 standard; Protein: 1417 AA.
XX
AC AAV31947;
XX
```

```
DT 21-DEC-1999 (first entry)
XX
XX Plasmodium falciparum cytoadherence gene protein ClAG3.1.
DE Plasmodium falciparum cytoadherence gene protein ClAG3.1.
XX
XX ClAG3.1; cytoadherence linked asexual gene; ClAG; erythrocyte;
KW red blood cell; malaria; infection; therapy; vaccine.
XX
OS Plasmodium falciparum.
XX
PN WO9949048-A1.
XX
PD 30-SEP-1999.
XX
PF 25-MAR-1999; 99WO-AU00213.
XX
PR 25-MAR-1998; 98AU-0002580.
XX
PA (MENZ-) MENZIES SCHOOL HEALTH RES.
XX
PI Kemp DJ, Trenholme KR, Gardiner DL, Holt DC, Cowman AF;
XX
DR WPI; 1999-591099/50.
XX
PT New proteins useful for treatment of Plasmodium infections in humans,
PT especially malaria
XX
PS Claim 16; Page 117-126; 150pp; English.
XX
CC This sequence represents the Plasmodium falciparum cytoadherence
CC linked asexual gene 3.1 (ClAG3.1) protein that facilitates
CC cytoadherence of erythrocytes parasitised with Plasmodium spp. to
CC other cells. It is encoded by a gene located on chromosome 3 of
CC P. falciparum. It is a member of a family of ClAG proteins (see
CC also AAV31945-49). The identification of clag genetic species, and
CC the products encoded by them, enables a range of therapeutic agents
CC to be rationally designed and/or identified that are useful for the
CC prophylaxis and treatment of disease conditions caused or
CC exacerbated by infection with Plasmodium spp., e.g. malaria,
CC especially in humans.
XX
SQ Sequence 1417 AA;
XX
Alignment Scores:
Pred. No.: 1 21e-10 length: 1417
Score: 211.00 Matches: 171
Percent Similarity: 34.55% Conservative: 114
Best Local Similarity: 20.73% Mismatches: 225
Query Match: 5.18% Indels: 315
DB: 20 Gaps: 46
US-09-727-892a-2 (1-2286) x AAV31947 (1-1417)
QY 115 TATATAACGTTACTATTCTGTAGCAATGCTGTTAATGCTATGAATTT----- 168
Db 300 TyrLysAsnTyrPheGluAlaLysSerIleLysPhePheSerIrrPolInLysIleGlu 319
QY 169 -----GATGTTGAAGTATTCGAGTTTCGATTCGATCTTTAT 204
Db 320 PheSerMetSerAspArgPheLysValLeuAspMetMetCysAspHisGluSerValTyr 339
QY 205 GAGCGATTTTATCGATATGGAAGAAGCGATACA-----ATCGAATAATGAAA 255
Db 340 -----TysSerGluLysLysArgArgLysThrTyrLeuLysValAspArgSerAsn 356
QY 256 ACAGATATTATCATGATTCGACATAACTGATATATAATACGATATCTTTTACTTAA 315
Db 357 ThrSerMet-----GluCysAsn-----IleLeuGlu 365
QY 316 GACACCATGCGTTATTTGATATATTACACGGGAATAATATATTAAATCTGCAGAA 375
Db 366 TyrLeuLeuHisTyrPheAsnLysTyrGlnLeuGlu-----IleIleLysThrThrGln 383
QY 376 GAAATGACACACACATTAAAA-----ATGAAGAAGGCTACTATT 414
```

[illegible]

Dd	729	MetAsnSerSerProIaGlyPheTyrPheSerAsnTyrIlnAsnPro-----	744
Qy	1105	ACTTTTACATGATGACCAATATATTTTCATTAATATACATTGATGAAGATGATTAAC	1164
Dd	745	-----TyrIleArgTyrSpleuhHisAsp	752
Qy	1165	GATGATTATTAAATTAATAAATTA-----TCACGTGATTACGT	1203
Dd	753	LysValLeuSerGlnLysPheGluProIleProLysMetAsnGlnIntrpSlnValLeuLys	772
Qy	1204	CAAAAGAT-----GTA	1215
Dd	773	SerLeuIleGluCysAlaTyrAspMetTyrPheGluGlnArgHisValLysAsnLeuTyr	792
Qy	1216	AAATACATGAATTAATGATATGATTACGTTAATATCATATCAAAATACATTAAAGATGAT	1275
Dd	793	LysTyrHisAsn-----LleTyrAsnIleAsn-----AsnLysLeuMetLeuMet	807
Qy	1276	CAAGCATTAAGGGATTGATTGATGCATACATACGTTAATTCGTT-----CTATA	1322
Dd	808	ArgAsp-----SerIleAsp-----LeuTyrLysAsnSlnPheSpAspAlleu	822
Qy	1330	TATGATGTGATACCTTTCATGACGATGATATATTTTCAAAACATATTTATTAACA	1389
Dd	823	PhePheIlaAspIlePheAsnMetArgLysTyrMetThrAlaThrProValTyrLys---	841
Qy	1390	CAAGTATGATTAATAAAACAAATCAATATACATCACCATTAGCATATAC-----	1440
Dd	842	-----LysValLysAspArgVal-----TyrHisThrLeuHis	852
Qy	1441	--ATTACT--GATGATATACAGCAACACCATCTACA-----	1473
Dd	853	SerIleThrGlnLysAsnSerValAsnPheTyrLysTyrGlyLleIleTyrGlyPheLysVal	872
Qy	1474	AATGAGAGGTTATGTTATCTAAAGTCGTTTAAAGCATTAATAGCGATA-----	1524
Dd	873	AsnLysGlnIleLeu-----LysGluValValAspGlnLeuTyrSerLleTyrAsnPhe	890
Qy	1525	-----CCGCAATTCAGTTCACATTTTAACCTATTC-----	1554
Dd	891	AsnThrAspIlePheThrAspThrSerPheLeuGlnThrValTyrLeuLeuPheArg	910
Qy	1555	-----CGTTTAGATGAT-----AACATGACATA	1578
Dd	911	IleGluGlnThrTyrArgThrGlnArgArgAspSpleLysIleSerValAsnValAlaPhe	930
Qy	1579	TAC--ATAATCATTAACCGGTAC-----AAAACATGCAACGATATTA-----	1620
Dd	931	PheMetAsnValAlaAsnAsnTyrSerLysLeuAsnLysGluGlnArgGlnIleGlnIle	950
Qy	1621	-----TTA	1623
Dd	951	HisAsnSerMetAlaSerArgTyrTyrAlaLysThrMetPheAlaAlaPheGlnMetLeu	970
Qy	1624	TTTCCTCATTTGTCACATACGT-----TCATTGTATTACTATATGGTT	1668
Dd	971	PheSerThrMetLeuSerPheAsnAsnValAspAsnLeuAspLysAlaTyrGlyLeuSerGln	990
Qy	1669	CCATTCCATTACTTAACGGAAGTGAATGACGACCAATTTATTTATTTGGCATCTGAT	1728
Dd	991	AsnIleGlnValAlaThrSerThrSerAlaPheLeuThrPheAlaTyrValTyrAsnGly	1010
Qy	1729	AGTTTGATATGAATCCGTTGTTAAACCCCTTATGACCCACGTTTATGACCCGATA	1788
Dd	1011	SerIle--MetAspSerValThrAsnSerLeuLeuProTyrAlaLysLysProIle	1029
Qy	1789	GCCTTAGCTAAATGGGATATTGAAAACGAACAGATGATAGATCTGTTGACTGATCAT	1848
Dd	1030	ThrIleuLysTyrGly---LysThrPheValPheSerAsnTyrPheMetLeuAlaSer	1048
Qy	1849	AAGAATAAT	1857
Dd	1049	LysMetTyr	1051

Thu Jan 9 14:50:04 2003

us-09-727-892a-2.rag

Page 21

Search completed: January 8, 2003, 17:09:01
Job time : 86.5 secs

This Page Blank (uspto)

Thu Jan 9 14:50:05 2003

us-09-727-892a-2.ra1

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: January 8, 2003, 16:14:14 (Search time 19.5 Seconds
(without alignments)
6898.544 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 4070
Sequence: 1 atggagatctagatgcacgtttaa2286

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09727892/rnat_06012003_151202_9347/app.query.fasta_1.2439
-DB=Issued-Patents_AA -QFMT=fasta -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORMEXT -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09727892.ecgn_1_11.ctunal.06012003_151202_9347 -NCPU=6 -ICPU=3
-NO_XLUPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONDILOG -DEV_TIMOUT=120
-WARN_TIMOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: Issued-Patents_AA:*
2: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/PCRTS.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	184	4.5	990	2	US-08-392-625-20
2	184	4.5	990	2	US-08-466-961A-20
3	183.5	4.5	2710	1	US-08-480-604A-6
4	183.5	4.5	2710	1	US-08-405-496A-6
5	183.5	4.5	2710	4	US-08-915-136-6
6	183.5	4.5	2710	4	US-08-957-310-6
7	183	4.5	3135	1	US-08-323-170B-2
8	181.5	4.5	3135	1	US-08-954-441-2
9	181.5	4.5	944	4	US-09-134-001C-4352
10	178.5	4.4	956	4	US-09-134-001C-4452
11	177.5	4.4	990	2	US-08-645-193B-15
12	174.5	4.3	1024	4	US-09-091-117-5

13	163.5	4.0	990	4	US-09-627-376-7	Sequence 7, Appli
14	162.5	4.0	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
15	161.5	4.0	2366	1	US-08-480-604A-10	Sequence 10, Appl
16	161.5	4.0	2366	2	US-08-405-496A-10	Sequence 10, Appl
17	161.5	4.0	2366	4	US-08-915-136-10	Sequence 10, Appl
18	161.5	4.0	2366	4	US-08-957-310-10	Sequence 10, Appl
19	159.5	3.9	648	4	US-09-370-861A-75	Sequence 75, Appl
20	158	3.9	677	3	US-08-480-604A-115	Sequence 115, App
21	158	3.9	677	3	US-08-480-604A-193	Sequence 193, App
22	158	3.9	677	3	US-08-295-802-115	Sequence 115, App
23	158	3.9	677	4	US-08-686-968C-58	Sequence 58, Appl
24	158	3.9	677	4	US-08-686-968C-193	Sequence 193, App
25	158	3.9	677	4	US-08-488-237A-115	Sequence 115, App
26	158	3.9	677	4	US-08-488-237A-193	Sequence 193, App
27	158	3.9	677	4	US-08-375-992A-115	Sequence 115, App
28	158	3.9	677	4	US-08-375-992A-193	Sequence 193, App
29	157	3.9	1226	2	US-08-540-804-12	Sequence 12, Appl
30	157	3.9	1226	2	US-08-218-265-12	Sequence 12, Appl
31	157	3.9	1226	3	US-08-521-872-12	Sequence 12, Appl
32	157	3.9	1226	4	US-08-590-399-12	Sequence 12, Appl
33	156.5	3.8	476	4	US-09-316-083-3	Sequence 3, Appli
34	156.5	3.8	993	4	US-08-836-687B-30	Sequence 30, Appli
35	155.5	3.8	3169	4	US-09-453-702B-257	Sequence 257, App
36	147.5	3.6	1296	1	US-08-480-604A-28	Sequence 28, Appl
37	147.5	3.6	1296	4	US-08-405-496A-28	Sequence 28, Appl
38	147.5	3.6	1296	4	US-08-915-136-28	Sequence 28, Appl
39	145	3.6	840	4	US-08-974-549A-190	Sequence 190, App
40	145	3.6	872	3	US-08-851-843A-8	Sequence 8, Appli
41	145	3.6	872	3	US-08-851-843A-54	Sequence 54, Appli
42	145	3.6	872	4	US-08-974-549A-221	Sequence 221, App
43	145	3.6	872	4	US-08-854-050-8	Sequence 8, Appli
44	145	3.6	872	4	US-08-854-050-54	Sequence 54, Appli
45	145	3.6	872	4	US-09-430-323-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-392-625-20
Sequence 20, Application US/08392625
Patent No. 5837485
GENERAL INFORMATION:
APPLICANT: Entian, Karl-Dieter
APPLICANT: Gtz, Friedrich
APPLICANT: Schnell, No. 5837485bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Gerhard
APPLICANT: Rosenstein, Ralf
APPLICANT: Kaletta, Cortina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G. nther
TITLE OF INVENTION: Biosynthetic Process For The Preparation
OF CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,625
FILING DATE:
CLASSIFICATION: 435

QY 1663 -----TTGGTTCT-----TTCCAACTTAACGAAAGTGAATTGACGAC--- 1704
Db 811 lIeValAspTyrValProGluValTyrGlyGlyProHisValIleGluAspIle 830
QY 1705 ---AATTATTATTGGGATGACT-----GATAGTTTGTATAG----- 1740
Db 831 GluAsnProPheMetTyrAspSerLeuAspSerIleAsnIleIleGlnSerGluPhe 850
QY 1741 -----AAATCCGTTGTAAACCTTATTGAACCCCAAGTTTATGACCCGATGCC 1791
Db 851 LysIleProLysGluPheIleValAlaIleSerIleAspPheLeuAspTyrLeuGlu 870
QY 1792 TTAGCTAAATGGATATTGAAAACGACAGATAGATGTTTGTACTGCAATCATAG 1851
Db 871 lIeAsnLys-----SerGluLysGluGluIle-----LeuIleAsnAsnAla 884
QY 1852 AAATATCATATGATGATGATGA-----AAGATTAAATATGCTTCTGCTGCT 1899
Db 885 GluAspLeuTyrTyrSerAsnAspIleArgGluTyrLysAsnLeuLeuAlaLysLeuThr 904
QY 1900 ATACCGAAAAAGCCCTTGATACAGCGTCGATTGTAACCTTTGACGTGA----- 1953
Db 905 AspProLysAsn-----AspTyrGluIleLeuLysGluPhePro 918
QY 1954 -----CAATCTTTGACGGTGGCATTTATGAAAACATTAAGATCTAT 1998
Db 919 AsnLeuHisGluPheLeuPheAsnLysIleSerIleLeuGluAsnLeuLysLysThrLeu 938
QY 1999 AATGAGCAAGGTACATATATCCGCTAAACGTAAATGTATGTATGTAATGTA 2058
Db 939 GlnLys-----SerLeuTyrThrSerArgSerArgIleIleGlnSerPheIle 954
QY 2059 TATGATGAATTTTACTGATGACACTAATATGAACGTGAATTTATTAATAAGACGCT 2118
Db 955 His-----MetArgCysAsnArgIlePheGlyIleAsnProGlu 967
QY 2119 AGACAAATTTGACACATAGTCAATTGATGATATT 2154
Db 968 LysGluLysPheValIleAspSerIlePheAsnGluIle 979

RESULT 2
US-08-466-961A-20
Sequence 20, Application US/08466961A
Patent No. 5843709
GENERAL INFORMATION:
APPLICANT: Entlan, Karl-Dieter
APPLICANT: G t z, Friedrich
APPLICANT: Schmeil, No. 5843709bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Gernar
APPLICANT: Rosenstein, Ralf
APPLICANT: Kaletta, Cortina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
TITLE OF INVENTION: Biosynthetic Process for the Preparation of
TITLE OF INVENTION: Chemical Compounds
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, NW
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,961A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/392,625
FILING DATE: 22-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/876,791
FILING DATE: 30-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/784,234
FILING DATE: 31-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.0980004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-466-961A-20

Alignment Scores:
Pred. No.: 7.73e-09 Length: 990
Score: 184.00 Matches: 152
Percent Similarity: 37.88% Conservative: 148
Best Local Similarity: 19.19% Mismatches: 286
Query Match: 4.52% Indels: 206
Gaps: 41

US-09-727-892A-2 (1-2286) x US-08-466-961A-20 (1-990)

QY 55 TATGTGGATTAACA---TTACGTACAAATAGTTACGACGACAAATACCAACC 111
Db 302 TyrLeuGlnIleAspThrTyrIleAspTyrSerArg-----AsnGluLeuLysGlnSer 319
QY 112 AAATATAAAACGTTACTATTCTGTGCAATTTGTTG---TTAATGTTATGAATT 168
Db 320 LeuAlaAspAsnIleSerGluAlaIleTyrIleLeuThrLeuLeuSerProHisGluPhe 339
QY 169 GATGTGAAGTATTTCCGAGTTTCGATCTTTTATGACGCAATTTATACGATGCAAA 228
Db 340 GlyThrLysThrIleArgAsnTyrHisGluPhePheMetAspLysTyrGlyPhe----- 357
QY 229 AGACGTATACATACAAATGCAAAATGCAAAACGATATTATCATGATTCACATACGTAT 288
Db 358 -----GluGlnLeuValAsnLeuLysGlnLeuLeuSerAspIleAsn 371
QY 289 AAATACGATATCATTTTATTACTTAAGACACCATCGTTATTGTAATATTAACA--- 345
Db 372 GlyPheGly-----TyrProLysLysAspSerTyrSerPheSerAsnIleAlaPhe 389
QY 346 CGGAAATATATATTAAATGTCGAGAAAGAAATGACACACATTAATAATGAAAGAG 405
Db 390 LeuLysGlnLysTyrLeuLeuAlaIleGlnAsnAsnSerHis---IleGluIleThrGlu 408
QY 406 GCTACTATT-----TTAGCCAAAATCAAAATGTAATTTTGAATAACGCTTAATCT 459
Db 409 AsnAspValLysAsnLeuGluLysAsnAsnThrVal-----SerLysIleAsnAla 425
QY 460 TCAATATTTAGATTTA-----ACAATGTTTAAATGTTTAA----- 501
Db 426 ProValSerThrGluIleTyrSerGluIleTyrPheGlyAsnSerIleLysGlyTyrGlu 445
QY 502 ---TTAATATTTATGATACATTTATGAACCAATACATCAATTCACATTAAGTAGAG 558
Db 446 AspPheAlaValIleSerProIleLeuGluLysSerPheAsnAlaGlyAlaThrPheGlyArg 465

QY 559 AATTAAGTGGTGGTTATTATTAACGAATCAACATTAAACAGATT-----AAT 612
 Db 466 ProThgIysPheasnIleuIysIysAsnGlnLeuGlnIleValHisHis 485
 QY 613 TATACATTTTGGATTAATAATATGATGATGATGATGATGATGATGATGATGATG 672
 Db 486 TyrAsnTyrMetAsnGlnAspLeuGlnIleSerGlnLeuAsnGln 502
 QY 673 AATGTTTGGCAAACTCACCTGACGACATACATACATACATACATACATACAT 732
 Db 503 -----AlaProLeuAsnSerArgAsnValAsnIleLeuAsnAsnAsnArgIleTyr 519
 QY 733 -----TTAGTATGTCGCATATCATATCATATCATATCATATCATATCA 771
 Db 520 AsnThrCysLeuAsnLeuAsnLeuProLysSerAspIleAspIleAsnIlePheIle 539
 QY 772 AATTTGACTATACAAATTAACATTTTTCATGATATATGAAATCTTACTTCAATAT 831
 Db 540 GlyAlaThrPheAsnLysLeu-----TyrLeuTyrSer 550
 QY 832 GAAATGACACGTTTCACTTACTCAACCAATATCAAGATTAATTAATATCTTATACACT 891
 Db 551 GluLys-----HisAspSerArgIleValIlePheValSer 561
 QY 892 TATCATTTCCATGATATGATGATTTTATGACTATATTAATCATTTGATGCTGCTTCA 951
 Db 562 AsnSerMetPheAsnTyrGlnPheGlySerGlnLeuTyrLysPheLeuArg-----GluIle 580
 QY 952 AATATGATTAACACCAATATACATTAACAACTAATGATGAGCGCTGTTTCTTATGAC 1011
 Db 581 SerPheGlnLysThrLysPheIleGlnProIleThrGlnGlu-----GlyIleAsp 597
 QY 1012 ATCAATTCGAGTATCCCTTATGATGATGATGATGATGATGATGATGATGATGATG 1062
 Db 598 SerLeuProPheCysProArgIleIleTyrLysAsnIleLeuLysProIleThrPro 617
 QY 1063 TTAATCTTTACGACACTATTCGACGACCAACGCTTAATCCCTTACTTTTACATGATGAC 1122
 Db 618 LysIleAsnSerGlnMetPheSerGlu-----ThrGln 628
 QY 1123 AATTATTTTCTATTATAT-----AAGATTGATTAAGATGA----- 1158
 Db 629 AsnThrLeuAsnArgPheIleThrIleArgLysThrPheIleProLysAspValIleIle 648
 QY 1159 -----TTTACGATGATTTTATTAATTAATTAATTAATTAATTAATTAAT 1185
 Db 649 AlaPheGlyAspAsnArgLeuLeuLeuAsnLeuLeuAsnLysHisLeuIleLeu 668
 QY 1186 AATATCAGCTATATACGTCAAATGATGTAATTAATTAATTAATTAATTAATTAATTAAT 1245
 Db 669 LysLysGlnLeuLysLysHisGlyArgIleArgIleLeuGlnSer-----PheIle 685
 QY 1246 AATATCAATACAAATACATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1302
 Db 686 Asn---GlnSerAsnAsnGlnArgMetLeuGlnIleValThrProLeuTyrLysLysThr 704
 QY 1303 CATATACGCTTATATCGCTTGTATATATGATGATGATGATGATGATGATGATGATGAT 1362
 Db 705 SerLeuLysGlnLeuLysHisGlyArgIleProLysAsnArgAsnLysHisPheAsnLeu 724
 QY 1363 -----ATTTTCAAACTATTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 1386
 Db 725 LysAspThrPheSerIleHisLeuSerIleProLysThrTyrGlnAsnProIleGln 744
 QY 1387 -----ACACAAGTAACTTAATAAAACAATAATTAATTAATTAATTAATTAATTAAT 1425
 Db 745 AspThrLeuLeuProPheIleThrGlnLeuLysValAsnAsnPheIleAsnLysPhePhe 764
 QY 1426 CTTTACAGATACATTAATTAATGATGATGATGATGATGATGATGATGATGATGATG 1485
 Db 765 TyrIleLysPheLysGlnAspGlnAspPheIleLysLeuThrGlnLeuArgGlnAsp---- 783
 QY 1486 ATCTTATCTAAAGTCGTTTAAATGATATATATGATACCTTCACTTACATTTT 1545

Db 783 ----- 783
 QY 1546 AACTTATCCGTTAGATGATTAACATGACATATACATATTAACGGTTACAAAAC 1605
 Db 784 -----GluAspTyrSerGlnIleTyrSerPheIleLysAsnThrLysAsp 798
 QY 1606 ACTGACGTAATATATATTTCTCTACATTTCTCACAACACCTTCACTGATTAATCTTA 1662
 Db 799 -----TyrCysLeuLeuAsnSerGlnLeuTyrAspTyrSer 810
 QY 1663 -----TTGGTTCCT-----TTCCAAATCTTAACGGAAGTGAATTAAGACAC 1704
 Db 811 IleValAspTyrValProGlnValTyrArgTyrGlyProHisValIleGluAspIle 830
 QY 1705 -----AATTATTTATTCGCACT-----CATGCTTGTATG 1740
 Db 831 GluAsnProPheMetTyrAspSerLeuLeuAspSerIleAsnIleIleGlnSerGlnPhe 850
 QY 1741 -----AATCGGTTGTTAAACCTTATTAACCCCACTTATTAACCCCACTTATTAAC 1791
 Db 851 LysIleProLysGlnPheIleValAlaIleSerIleAspPheLeuLeuAspTyrLeuGln 870
 QY 1792 TTAGCTAAATGGATATTTGAACGACATGATTAAGATGTTTGTACTGATATCAATAG 1851
 Db 871 IleAsnLys-----SerGlnLysGlnGlnIle-----LeuIleAsnAsnAla 884
 QY 1852 AATATGCAATPAAGTGAATGGA-----AAGATTAAATTTGCTTCTGCTGT 1899
 Db 885 GluAspLeuTyrArgSerAsnAspIleArgGlnTyrLysAsnLeuLeuAlaLysLeuThr 904
 QY 1900 ATACCGAAAAAGCGCTTGATPAACAAGCGTGAATTTGAACCTTTGTACGAA----- 1953
 Db 905 AsnProLysAsn-----AspTyrGlnIleLeuLysLysGlnPhePro 918
 QY 1954 -----CAATCTTGACGTCGCTTATTTGAAACATTAATTAATTAATTAATTAATTAAT 1998
 Db 919 AsnLeuHisGlnPheLeuPheAsnLysIleSerIleLeuGlnAsnLeuLysLysThrLeu 938
 QY 1999 AATGACGAGGTACAAATATGATATATCCGTCTAAACGTAATGATGATGATGATGATGAT 2058
 Db 939 GlnLys-----SerLeuTyrThrSerArgSerArgIleIleGlySerPheIle 954
 QY 2059 TATGATGAATATTTACTGATGACACTTAATATGAACGCTGAATTTATATTAAGACGCT 2118
 Db 955 His-----MetArgCysAsnArgIlePheGlyIleAsnProGln 967
 QY 2119 AGAGAAATTTGACCATGATCAATTTGATGATATTT 2154
 Db 968 LysGlnLysPheValLeuSerIlePheAsnGlnIle 979

RESULT 3
 US-08-480-604A-6
 : Sequence 6, Application US/08480604A
 : Patent No. 5736139
 : GENERAL INFORMATION:
 : APPLICANT: KINK, JOHN A.
 : APPLICANT: THALLEY, BRUCE S.
 : APPLICANT: PADHIE, NISHA V.
 : APPLICANT: FIRCO, JOSEPH R.
 : APPLICANT: STAFFORD, DOUGLAS C.
 : TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
 : PREVENTION OF C. DIFFICILE DISEASE
 : NUMBER OF SEQUENCES: 32
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: MEDLEN & CARROLL, LLP
 : STREET: 220 MONTCOMERY STREET, SUITE 2200
 : CITY: SAN FRANCISCO
 : STATE: CALIFORNIA
 : COUNTRY: UNITED STATES OF AMERICA
 : ZIP: 94104
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Alignment Scores:
Pred. No.: 1.19e-08
Score: 183.50
Percent Similarity: 34.66%
Best Local Similarity: 19.55%
Query Match: 4.51%
DB: 1
Gaps: 46

US-09-727-892a-2 (1-2286) x US-08-480-604A-6 (1-2710)
QY 175 GAAGTATTCGAGT-----TTGCAATCTTTTATGACGATTTATAGC 219
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1258 AspleuTyrrProclLyshPheTyrrTPaGpHeTyrrAlahePheSpTyrrAlaIleThr 1277
QY 220 TAATGTGAAA-----AGACGTGATACACATCACAATAACAAAACAGAT----- 261
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1278 ThrLeuTySerProValTyrrGluAspThrAsnIleLysIleLysLeuAspLysAspThrArg 1297
QY 262 ---ATTATCATGTGATGCACTACCTGAATTAATACGATATATCTTTTACTTAAGAC 318
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1298 AsnPhelIleMetProThrIleThrThrAsnGlu-----IleArgAsn 1311
QY 319 ACCATGCGTTAT---TTTGAT----- 336
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1312 LysLeuSerTyrrSerPheAspGlyAlaGlyGlyThrTyrrSerLeuLeuLeuSerSerTyrr 1331
QY 337 -----AATATTACACCGGAATAATATATATTATAAATCTGCAGAGAGA 378
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1332 ProIleSerThrAsnIleAsnLeuSerLysAspAspLeuThrPhePheAsnIleAspAsn 1351
QY 379 AATGAACACACATTAAATAAAGAAGAGCGTACTATT----- 414
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1352 GluValAlaGluIleSerIleGluAsnGlyThrIleLysLysGlyLysLeuIleLysAsp 1371

```

```

QY 415 -----TTAGCCAAAATCAATGTAAT----- 438
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1372 ValLeuSerLysIleAspIleAsnLysAsnLysLeuIleIleLysAsnGlnThrIleAsp 1391
QY 439 -----TTAGCAAAA 447
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1392 PheSerGlyAspIleAspAsnLysAspArgTyrrIlePheLeuThrCysGluLeuAspAsp 1411
QY 448 CGGTGTAATCTTCATTCATTTAGATTTA-----ACATGTTTTTAAT 492
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1412 LysIleSerLeuIleIleGluIleAsnLeuValAlaLysSerTyrrSerLeuLeuSer 1431
QY 493 GGTTTTAATTTTAAATTTATGTAATTTGTAAGAAACCAATACATCAATTCGACATTA 552
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1432 GlyAspLysAsnTyrrLeuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451
QY 553 GGT-----AGCAATTACTTGATGGTGGTTATTTAACAGAAATCACA----- 594
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1452 GlyLeuAspSerLysAsnIleAlaTyrrAsnTyrrThrAspGluSerAsnAsnLysTyrrPhe 1471
QY 595 -----CTTAAACAGATTTTAAATTATAGATTTTGCATTAAGATATAGATATGAT 645
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHisTyrrLysLysAspSerLysAsn 1491
QY 646 GATAGTGAAGCCTATGACTATGCTGGAATGTTTGCAAAACCTCACACCTGACAACTT 705
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1492 IleLeuGluPheTyrrAsnAspSerThrLeuGluPhe-----AsnSerLys 1506
QY 706 ACTATACATTCATATACCTGATTAATATAGATAGTGCATATTCATATATAGTATATA 765
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1507 AspPheIleAlaGluAspIle-----AsnVal 1515
QY 766 TTTCCAATTTTGACTATACAAATTAACA----- 795
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1516 PheMetLysAspAspIleAsnThrIleThrGlyLysTyrrValAspAsnThrAsp 1535
QY 796 -----TTTTCATGGAATATATGAAATCT----- 819
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1536 LysSerIleAspPheSerIleSerLeuValSerLysAsnGluValLysValAsnGlyLeu 1555
QY 820 TACTTGATTAAGAATATGACACGTTTTCAGTACCAACCAATCAACAAAT-----ATTA 876
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1556 TyrrLeuAsnGlu-----SerValTyrrSerTyrrLeuAspPheValLys 1570
QY 877 ATATCTTATACACATTAATCATTTCCATGAT---ATGAAT---TTTATGACTATATTA 930
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1571 AsnSerAspGlyHisHisAsnThrSerAsnPheMetAsnLeuPheLeuAspAsnIle 1589
QY 931 TCATTTCTATCGTGCGTTAAATATGTAATACACCAATATACATAACCAACTATGAT 990
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1590 SerPheTrpLys-----LeuPheGlyPheGluAsnIleAsnIleAsnValIleAsp 1605
QY 991 GAOCCTGTTTTCATATGACATCAATTCGAGTTATCCATAGTATGATATGAGAAA 1050
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1606 Lys---TyrrPheThrLeuValGlyLysThrAsnLeuGlyTyrrVal----- 1619
QY 1051 ATTCACACATGGTTATACCTTTACGACACCTATTGAGAACCAAGCTTAATCCCTCTTT 1110
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1620 -----GluPhe 1621
QY 1111 TTAGATGATGACAT---TATTTTCATATATATAGATGATGATAAGAT 1155
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1622 IleCysAspAsnAsnLysAsnIleAspIleTyrrPheGlyGluTrpLysThrSerSer 1641
QY 1156 GTATTTAACGATGATTATTAATTAATAATTAATCACTGCTTTA-----CGT 1203
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1642 -----LysSerThrIlePheSerGlyAsnGlyArg 1651
QY 1204 CAATGATGTAAATATATATATATATGATATGATATGATTCGTTATATCAATCAAT 1263
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1652 AsnValValValGluProIleTyrrAsnProAspThrGlyGluAspIleSerThrSerLeu 1671

```

QY 1264 TTAAGATATGATTCACACATACGCGTATTTGAT----- 1296
 Db 1672 AspPheSerTyrGluProLeuTyrGlyIleAspArgTyrIleAsnLysValLeuIleAla 1691
 QY 1297 -----TGCATGCATATACGCTGTTAAATCGTTGTTAATATGAAATG 1338
 Db 1692 ProAspLeuTyrThrSerLeuIleAsnIleAsnThrAsn-----TyrTyrSerAsn 1708
 QY 1339 GAATACTTCATGACAGTGTATTTTTCATAAACTATTTTATTAATAACACAGGTAAAG 1398
 Db 1709 GluTyrTyrProGluIleLeuValLeuAsnProAsnThrPheHisLys----- 1724
 QY 1399 TTAATAAACAAATCAATATG---ACATCACCTTACGACATACATTCATCGATGAT--- 1452
 Db 1725 ---LysValAsnIleAsnLeuAsnSerSerPheGluTyrLysTyrSerThrGluGly 1743
 QY 1453 -----ATCAACGACACCCCATCTCAATATGAGAGGTTATGTTATCTAAA 1497
 Db 1744 SerAspPheIleLeuValArgTyrLeuGluGluGluSerAsnLysIle---LeuGluIle 1762
 QY 1498 GTCGTTTAAATGATTTAATGCGATACCTGCATTCACATTCATTCATTCATTCGCT 1557
 Db 1763 IleArgIleLys-----GlyIleLeuSerAsnThrGlnSerPheAsnLysMetSer 1779
 QY 1558 TTAGATGATACAGACATGACATATACATATTAACGGTTACAAAACACTGAACTAAT 1617
 Db 1780 IleAsp-----PheLysAspIleLysLysLeuSerLeuGly 1791
 QY 1618 ATTTTATTTCTACATTTGTGCACATCAGCTTATGTTATTAATGTTGTTCTTCCAA 1677
 Db 1792 TyrIleMetSerAsnPheAsnSerPheAsnSerGluAsnGluLeu-----AspArgAsp 1809
 QY 1678 TACTTAACGGAAGTGAATTAATGACGACAAATTTATTTATGCGATACGATAGTTGTAT 1737
 Db 1810 HisLeuGluPheLysIleLeuAspAsnLysThrTyrTyrAspGluAspSerIleu 1829
 QY 1738 ATGAATTCGGTGTGTTAAACCTTATTGAACCCAGTTTA-----TTGCACCCGATAGCC 1791
 Db 1830 ValLysGlyLeuIleAsn-----IleAsnAsnSerLeuPheTyrPheAsnPro----- 1845
 QY 1792 TTACGTAATGCGATTTGTAACACGACGATGATGATGATGTTGTAATGATCATAG 1851
 Db 1846 -----IleGluPheAsnLeuValThrGlyTyrGluIleThrIleAsnGlyLys 1860
 QY 1852 AAATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1902
 Db 1861 LysTyrTyrPheAspIleAsnThrGlyAlaIleLeuThrSerTyrLysIleIleAsnGly 1880
 QY 1902 ----- 1902
 Db 1881 LysHisPheTyrPheAsnAsnAspGlyValMetGluLeuGlyValPheLysGlyProAsp 1900
 QY 1903 -----CCGAAAACGCGCTTTGATACAGCGTGCATTTTGAACCTTT 1944
 Db 1901 GlyPheGluTyrPheAlaProAlaAsnThrGlnAsnAsnAsnIleGluIleAlaIle 1920
 QY 1945 GTA-----CGTGAACAATTCCTTTACCGT----- 1968
 Db 1921 ValTyrGlnSerLysPheLeuThrLeuAsnGlyLysLysTyrTyrPheAspAsnAsnSer 1940
 QY 1969 -----CCCATTTATGAACAAATTAAGTATCATATGATGACCAAGCT 2010
 Db 1941 LysAlaValThrGlyTyrPargIleIleAsnAsnGluLysTyrPheAsnProAsnAsn 1960
 QY 2011 ACAATTCGATATATCGCTCAAACTGAATGTATGTGTATGATATGATGATATAT 2070
 Db 1961 AlaIleAlaIleAlaVal-----GlyLeuGluValIleAspAsnAsnLysTyr-----Tyr 1976
 QY 2071 TTATACGATGAGA-----CTTAATATGAAGCTGGA 2100
 Db 1977 PheAsnProAspPheAlaIleIleSerLysGlyTyrPglThrValAsnLysSerArgTyr 1996
 QY 2101 TTATATTTAAAGACGCTAGAGAAATTTGACCATAGTCAATTGATGAT-----ATT 2154

Db 1997 TyrPheAspThrAspThrAlaIleAlaPheAsnGlyTyrIleAsnGlyLysHis 2016
 QY 2155 CTTTATATGGAAGTGC-----ATCGGTTCTATTTCATCTTAACGACTTATTT 2202
 Db 2017 PheTyrPheAspSerAspCysValIleLysIleGlyValPheSerThrSerAsnGlyPhe 2036
 RESULT 4
 US-08-405-496A-6
 : Sequence 6, Application US/08405496A
 : Patent No. 5919665
 : GENERAL INFORMATION:
 : APPLICANT: WILLIAMS, JAMES A.
 : TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
 : TITLE OF INVENTION: NEUROTOXIN
 : NUMBER OF SEQUENCES: 30
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: MEDLEN & CARROLL, LLP
 : STREET: 220 MONTGOMERY STREET, SUITE 2200
 : CITY: SAN FRANCISCO
 : STATE: CALIFORNIA
 : COUNTRY: USA
 : ZIP: 94104
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/405,496A
 : FILING DATE: 16-MAR-1995
 : CLASSIFICATION: 424
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/329,154
 : FILING DATE: 25-OCT-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/161,907
 : FILING DATE: 02-DEC-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/985,321
 : FILING DATE: 04-DEC-1992
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/429,791
 : FILING DATE: 31-OCT-1989
 : ATTORNEY/AGENT INFORMATION:
 : NAME: INGOLIA, DIANE E.
 : REGISTRATION NUMBER: 40,027
 : REFERENCE/DOCKET NUMBER: OPHD-01308
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 705-8410
 : TELEFAX: (415) 397-8338
 : INFORMATION FOR SEQ ID NO: 6:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2710 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-405-496A-6
 Alignment Scores:
 Pred. No.: 1.19e-08 Length: 2710
 Score: 183.50 Matches: 172
 Percent Similarity: 34.66% Conservative: 133
 Best Local Similarity: 19.55% Mismatches: 270
 Query Match: 4.51% Indels: 305
 DB: 2 Gaps: 46
 US-09-727-892A-2 (1-2286) x US-08-405-496A-6 (1-2710)
 QY 175 GAAGTATTCGAGT-----TTGCAATCTTTTGAAGCATTTTATACG 219
 Db 1258 AspLeuTyrProGlyLysPheTyrTyrPargPheTyrAlaPhePheAspTyrAlaIleThr 1277


```

Db 1901 GlyPheGluTyrPheAlaProAlaAsnThrGlnAsnAsnIleGluGlnAlaIle 1920
QY 1945 GTA-----CGTGAACAATTCTGTCACGCT----- 1968
    |||         |||         |||         |||         |||
Db 1921 ValTyrGlnSerLysPheLeuThrLeuAsnGlyLysLysTyrTyrPheSpasnsnSer 1940
QY 1969 -----GCCATTATTTGAAACATAAAGATCATCATATGACGACAGT 2010
    |||         |||         |||         |||         |||
Db 1941 LysAlaValThrGlyTyrPargIleLeuAsnGlyLysTyrTyrPheAsnProAsnAsn 1960
QY 2011 ACAATATCGATATATCCGCTAAACTGAATGTATGTGTAATGATATGATATAT 2070
    |||         |||         |||         |||         |||
Db 1961 AlaIleAlaAlaVal-----GlyLeuGlnValIleAspAsnLysTyr-----Tyr 1976
    |||         |||         |||         |||         |||
QY 2071 TTATCTGATGAA-----CTTAATATGAAACGCTGAA 2100
    |||         |||         |||         |||         |||
Db 1977 PheAsnProAspThrAlaIleIleSerLysGlyTyrPglThrValAsnGlySerArgTyr 1996
QY 2101 TTATATTTAAACACCGCTACAGCAAAATTCGACCATGACATTTGATAT-----ATT 2154
    |||         |||         |||         |||         |||
Db 1997 TyrPheAspThrAspThrAlaIleAlaPheAsnGlyTyrLysThrIleAspGlyLysHis 2016
QY 2155 CTTTATATGTAAGTGAC-----ATCGGTTTCACTTTCACCTTACGACTTATTT 2202
    |||         |||         |||         |||         |||
Db 2017 PheTyrPheAspSerAspCysValValLysIleGlyValPheSerThrSerAsnGlyPhe 2036

RESULT 5
US-08-915-136-6
Sequence 6, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADRYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989

```

```

ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPBD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6

Alignment Scores:
Pred. No.: 1.19e-08 Length: 2710
Score: 183.50 Matches: 172
Percent Similarity: 34.66% Conservative: 133
Best Local Similarity: 19.55% Mismatches: 270
Query Match: 4.51% Indels: 305
DB: Gaps: 46

US-09-727-892A-2 (1-2286) x US-08-915-136-6 (1-2710)
QY 175 GAAGTATTTCCGAGT-----TTGCAACTTTTATGACGCAATTTTATACG 219
    |||         |||         |||         |||         |||
Db 1258 AspleuTyrProGlyLysPheTyrTyrPargPheTyrAlaPheAspTyrAlaIleThr 1277
QY 220 TATGTGAAA-----AGACGTGATACATACACAAATCAAAACAGAT----- 261
    |||         |||         |||         |||         |||
Db 1278 ThrLeuLysProValTyrGlyAspThrAsnIleLysIleLysLeuAspLysAspThrArg 1297
QY 262 ---ATTATCATGATTCGACATACACTGATATATAATACGATATATCTTTTACTTAAGAC 318
    |||         |||         |||         |||         |||
Db 1298 AsnPhelIleMetProThrIleThrThrAsnGlu-----LleArgAsn 1311
QY 319 ACCATGCGCTTAT-----TTTGAT----- 336
    |||         |||         |||         |||         |||
Db 1312 LysLeuSerTyrSerPheAspGlyAlaGlyLysThrTyrSerLeuLeuSerSerTyr 1331
QY 337 -----AATATTACACGCGAAATATATATTATTAATCTGCAGAAAGAA 378
    |||         |||         |||         |||         |||
Db 1332 ProlSerThrAsnIleAsnLeuSerLysAspSpleuThrIlePheAsnIleAspAsn 1351
QY 379 AATGAACACACATTAATAAATGAAGAGCGTACTATTT----- 414
    |||         |||         |||         |||         |||
Db 1352 GluValArgGluIleSerIleGluAsnGlyThrIleLysLysGlyLysLeuIleLysAsp 1371
QY 415 -----TTAGCCAAAATCAAAATGTATTT----- 438
    |||         |||         |||         |||         |||
Db 1372 ValLeuSerLysIleAspIleAsnLysAsnLysLeuIleIleGlyAsnGlnThrIleAsp 1391
QY 439 -----TTAGAAA 447
    |||         |||         |||         |||         |||
Db 1392 PheSerGlyAspIleAspAsnLysAspArgTyrIlePheLeuThrCysGluLeuAspAsp 1411
QY 448 CGTGTAAATCTTCATCATCAATTTAGATTTA-----ACAATGTTTAAAT 492
    |||         |||         |||         |||         |||
Db 1412 LysIleSerLeuIleIleGluIleAsnLeuValAlaLysSerTyrSerLeuLeuSer 1431
QY 493 GGTTTAAATTTATATTATGATPACTTTATGAAAACCAATACATCAATTTGCAACATTA 552
    |||         |||         |||         |||         |||
Db 1432 GlyAspLysAsnTyrLeuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451
QY 553 GGT-----AAGAAATCTTGATGCTGATGTTATTAACAGATCAACA----- 594
    |||         |||         |||         |||         |||
Db 1452 GlyLeuAspSerLysAsnIleAlaTyrSerTyrThrAspGluSerAsnAsnLysTyrPhe 1471
QY 595 -----CTTAACACAGATTTAATTAATATACGATTTTGTATAAGATTAATGATATGAT 645
    |||         |||         |||         |||         |||
Db 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHisTyrLysLysAspSerLysAsn 1491

```

QY	646	GATGATGAACCCATGCTATGCTGTGGAAATGTTTGGCAAAAGCTCACACCGTAACAACTT	705
Db	1492	ILEUGSLHERFYRSLSPSSPTNRLLEUGLPHPE-----ASNSRLYS	1506
QY	706	ACATACATTCATATGACGCTGATTTATATAGGTATGCGCAATTCATTAATAGTGATATA	765
Db	1507	ASPHETLEALGLAUNSPLE-----ASNVAL	1515
QY	766	TTTCCAAATTTTGACTATACAAATTAA--	795
Db	1516	PHMETLSPSPSLSPLEASNTNRLLETNRLGLYSYTYRVALSPASASNTHRASP	1535
QY	796	-----TTTCATCGAATTTATATGCAATC-----	819
Db	1536	LYSSERLEASPHESERLEISERLEUVALSERLYSASGNIVALYSVALSNGLYEU	1555
QY	820	TACTTCATATATGAATGACACGGTTTTCAGTTACTCAACCAATATCANGAT---ATTAA	876
Db	1556	TYRLEUASNGLU-----SERVALTYRSERTYRLEUASPHEVALYS	1570
QY	877	ATATCTATACATATATCATTTCCATGAT---ANGAT---TTTATGACCATATTA--	930
Db	1571	ASNSERASPOLYHSHISASNTSRASNPHEWASLEUPHEUASPSNTLE---1589	
QY	931	TCATCTATCGTGGTGGTTAAATATATGATTAACACCAACCAATACAAACCAATATGAT	990
Db	1590	SERPHETRPLYS-----LEUPHEGLYPHEGLUASNLEASNPHEVALILEKSP	1605
QY	991	GAGCTGTGTTTCTATGTGACATCATTTGCGAGTATCCGTATGCTATGATCAIGAAAA	1050
Db	1606	LYS---TYRPHETRLEUVALGLYLSYSTRASLEUGLYTYVAL-----	1619
QY	1051	ATTCACAATGTTAACTTTTACGACACTTTCAGAACCAACGTTATCCCTACTTTT	1110
Db	1620	-----GLUPHE	1621
QY	1111	TTAATATACATC-----TATTTTCTATTAATAGATTGATAAGAT	1155
Db	1622	ILECYSASPASNSLYSASNLEASPILETYRPHGLYGLUTRPLYSRHSERSESR	1641
QY	1156	GTATTTACCATGATTTATTAATTAATAATTAATACACGCTATTA-----CGT	1203
Db	1642	-----LYSERTRHILEPHERSEGLYASNDLYARG	1651
QY	1204	CAATGATGTGTAAATACATATTAATAATGATATGATTTACGTTAATATCAATACAAATACA	1263
Db	1652	ASNVALVALGLUPROLETYRASNPROASPTHGLYGLUASPLSESRTHRSERLEU	1671
QY	1264	TTAGATATGATTCAGACATTTAGGGTATGAT-----	1266
Db	1672	ASPHSESRTYRGLUPROLETYRGLYLEASPAQTYRILEASNLYSVALLEULEALA	1691
QY	1297	-----TGCAAGCAWATACGGTTAAATCGTGTGTTATATATGAAATGT	1338
Db	1692	PROASPREUTYRTHRSERLEULEASNLEASNTHRASN-----TYTRYSERASN	1708
QY	1339	GAAATCTTTGATGACGCTGATATTTATTTTCAAACTATTTTATTAATAACACACAGCTAAG	1398
Db	1709	GLUTYRTRPROGLUILELEVALLEUASNPASNTNHPHENHLYS-----	1724
QY	1399	TTAAATAACAAATCATATG---ACATACACCTTACGATACATACATCATGATAT---	1452
Db	1725	---LYSVALSNLEASNTLEUASPSESRSESRPHGLUTYRLYSTYRSESRTHRICUDLY	1743
QY	1453	-----ATCAACGACACCAACCATCTCAATGAGGAGCGTTATGTTATCTAAA	1497
Db	1744	SERASPHETLEUVALARGTYRLEUGLUGLISERASNLYSVALSLE---LEUANDLYS	1762
QY	1498	GTCCTTTTAAATGATATATATGCAATACCGCATTCAGTTACATTTATTAATTCGCT	1557
Db	1763	ILEARGILEYS-----GLYLEUENSRASNTNHPGINSERPHENSNLYSSESR	1779
QY	1558	TTAATGATATAACGATACACTATACATATCATTTACGTTTACAAAAACGATGACGTAT	1617

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPED-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-6

Alignment Scores:
Pred. No.: 1,19e-08 Length: 2710
Score: 163.50 Matches: 172
Percent Similarity: 34.668 Conservative: 133
Best Local Similarity: 19.55% Mismatches: 270
Query Match: 4.51% Indels: 305
Gaps: 46

US-09-727-892a-2 (1-2286) x US-08-957-310-6 (1-2710)
QY 175 GAAGTATTCGAGT-----TTGGAATCTTTTATGACGCATTTTATGAC 219
Db 1258 ASPLenlyrProGlyysPheTyrrTParqPheTyrrAlaPhePheSPtyrAlaIleThr 1277
QY 220 TATGTGAA-----AGACGTGATACATCACAAATCAAAACAGAT----- 261
Db 1278 ThrLeuLyseroValTyrrGluAspThrAsnIleLysIleLysLeuAspLysAspThrArg 1297
QY 262 ---ATTATCATGATGCACATACTGTAATAATACGATATCATTTTACTTAAAGAC 318
Db 1298 AsnPheIleMetProThrIleThrIleThrAsnGlu-----IleArgAsn 1311
QY 319 ACCATGGCTAT---TTTGAT----- 336
Db 1312 LysLeuSerTyrrPheAspGlyAlaGlyGlyThTyrrSerLeuLeuLeuSerSerTyrr 1331
QY 337 -----AATATTACGCGGAAATATATATTAAATCGCAGAAAGAA 378
Db 1332 ProIleSerThrAsnIleAsnLeuSerLysAspAspLeuTriplPheAsnIleAspAsn 1351
QY 379 AATGACACACATTAATAATGAAGAGCTACATT----- 414
Db 1352 GluValArgGluIleSerIleGluAsnGlyThrIleLysLysGlyLysLeuIleLysAsp 1371
QY 415 -----TTAGCCAAATAATCAAAATGTAAT----- 438
Db 1372 ValLeuSerTyrrIleAspIleAsnLysAsnLysLeuIleIleGlyAsnGlnThrIleAsp 1391

```

```

QY 439 -----TTAGAAAA 447
Db 1392 PheSerGlyAspIleAspAsnLysAspArgTyrrIlePheLeuThrcyGluLeuAspAsp 1411
QY 448 CGTGTAAATCTTCATCATTTAGATTGA-----ACATGTTTAAAT 492
Db 1412 LysIleSerLeuIleIleGluIleAsnLeuValAlaLysSerTyrrSerLeuLeuSer 1431
QY 493 GGTTTAAATTAATATATATGATTACTTTAGAAACCAATACATCATATGCAACATTA 552
Db 1432 GlyAspLysAsnTyrrLeuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451
QY 553 GGT-----AAGAAATTAAGTATGATGATGATTAATTAACAGATCAACA----- 594
Db 1452 GlyLeuAspSerLysAsnIleAlaTyrrAsnTyrrThrAspGluSerAsnLysTyrrPhe 1471
QY 595 -----CTTAAACACATTTTAAATATACGATTTTGTAAAGATATGATGAT 645
Db 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHisTyrrLysAspSerLysAsn 1491
QY 646 GATAGTGAACCTATGACATAGCTGTGAATGTTTGCAGAAACACACCTGAACAACCTT 705
Db 1492 IleLeuGluPheTyrrAsnAspSerThrLeuGluPhe-----AsnSerLys 1506
QY 706 ACATACATTCATATAGACGTGATTAATAGGATATGACATATTCATTAATGATGATTA 765
Db 1507 AspPheIleAlaGluAspIle-----AsnVal 1515
QY 766 TTTCAGAAATTTGACTATACAAATTACA----- 795
Db 1516 PheMetLysAspAspIleAsnThrIleThrGlyLysTyrrValAspAsnAsnThrAsp 1535
QY 796 -----TTTTCATGATATATATGGAATCT----- 819
Db 1536 LysSerIleAspPheSerIleSerLeuValSerLysAsnGluValLysValAsnGlyLeu 1555
QY 820 TACTTGAAATTAAGATATGACACGTTTCACTTACACCAATATTCAGAT---ATTA 876
Db 1556 TyrrLeuAsnGlu-----SerValTyrrSerTyrrLeuAspPheValIleAsp 1570
QY 877 ATATCTTATACCATATATTCATGAT---ATGAT---TTTATGATATATTAATA 930
Db 1571 AsnSerAspGlyHisHisAsnThrSerAsnPheMetAsnLeuPheLeuAspAsnIle 1589
QY 931 TCATTCCTATCGGTGGTGTAAATATGTATATACACCAATACATTAACCAATATGAT 990
Db 1590 SerPheTrpLys-----LeuPheGlyPheGluAsnIleAsnPheValIleAsp 1605
QY 991 GAGCCTGTTTTCATTCATGACATCAATCGAGTATCTTATGTGATGATCATGAGAA 1050
Db 1606 Lys---TyrrPheThrLeuValGlyLysThrAsnLeuGlyTyrrVal----- 1619
QY 1051 ATTCACATGGTTATTAATTTTACGACACATATTCAGACACGATTAATCCCTACTTT 1110
Db 1620 -----GluPhe 1621
QY 1111 TTAGATGATGACAT-----TATTTCATTAATATAGATGATGATAAGAT 1155
Db 1622 IleCysAspAsnAsnLysAsnIleAspIleTyrrPheGlyGluTrpLysThrSerSer 1641
QY 1156 GTATTTAACGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1203
Db 1642 -----LysSerThrIlePheSerGlyAsnGlyArg 1651
QY 1204 CAATGATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1263
Db 1652 AsnValValValGluProIleTyrrAsnProAspThrGlyGluAspIleSerThrSerLeu 1671
QY 1264 TTAAGAAATGATTAACAGACATTAACGGGATGATGAT----- 1296
Db 1672 AspPheSerTyrrGluProLeuTyrrGlyLysAspArgTyrrIleAsnLysValLeuIleAla 1691
QY 1297 -----TGCATGATATACGCTTAATTTGTTTATTAATGAATGT 1338

```



```

Db 1692 ProAspLeuTyrThrSerLeuIleAsnIleAsnThrAsn-----TyrTyrSerAsn 1708
      :::::|||||
QY 1339 GAATACCTTCATGCACGCTGATATATTTTCAAAACTATTTTATAAACAAGCTAAG 1398
      |||||:::
Db 1709 GluTyrTyrProGluIleIleValIleAsnProAsnThrPheHisLys----- 1724
      |||||:::
QY 1399 TTAATAAACAATCATATG---ACATCAGCTTACGACATCTCATTTACTGATGAT--- 1452
      |||||:::
Db 1725 ---LysValAsnIleAsnLeuAspSerSerPheGluTyrLysTrpSerThrGluGly 1743
      |||||:::
QY 1453 -----ATCAAGACACCACTACCAATAGGAGGAGGTATGTTACTTAAA 1497
      :::::|||||
Db 1744 SerAspPheIleLeuValAlaGlyTyrLeuGluIleSerAsnLysLysIle---LeuGluLys 1762
      :::::|||||
QY 1498 GTCTGTTTAAATGATATATATGACATCTGATACCTTACATTTTAACTTATTCGCT 1557
      :::::|||||
Db 1763 IleArgIleLys-----GlyIleLeuSerAsnThrGlnSerPheAsnLysMetSer 1779
      :::::|||||
QY 1558 TTAGATGATACATGACACTTACATATATCATTAACGTTACAAACACGTAACGTAAT 1617
      :::::|||||
Db 1780 IleAsp-----PheLysAspIleLysLysLeuSerLeuGly 1791
      :::::|||||
QY 1618 ATATTATTCCTCATTTTGCACATCAGCTTCATTTGATATTAACCTTCTTCCAA 1677
      :::::|||||
Db 1792 TyrIleMetSerAsnPheLysSerPheAsnSerGluAsnGluLeu-----AspArgAsp 1809
      :::::|||||
QY 1678 TACTTACAGGAAGTGAATGACAGACATTTTATTTTCCGATCTGATAGTTTGTAT 1737
      :::::|||||
Db 1810 HisLeuGlyPheLysIleIleAspAsnLysThrTyrTyrTyrAspGluAspSerLysLeu 1829
      :::::|||||
QY 1738 ATGAATCCGTTGTTAAACCTTATTTGACCCAGCTTA-----TTGACCCGATGCC 1791
      :::::|||||
Db 1830 ValLysGlyLeuIleAsn-----IleAsnAsnSerLeuPheTyrPheAspPro----- 1845
      :::::|||||
QY 1792 TTAGTAAATGGATATTGAAAACGACAGATGAATGAATGTTTGTACTGATCATAG 1851
      :::::|||||
Db 1846 -----IleGluPheAsnLeuValThrGlyTyrGlnThrIleAsnGlyLys 1860
      :::::|||||
QY 1852 AAATATGCATATGAAGTGAAGTGAAGATTAATTTGCTGCTGCTGATAT----- 1902
      :::::|||||
Db 1861 LysTyrTyrPheAspIleAsnThrGlyAlaIleLeuThrSerTyrLysIleIleAsnGly 1880
      :::::|||||
QY 1902 ----- 1902
      :::::|||||
Db 1881 LysHisPheTyrPheAsnAsnAspGlyValMetGlnLeuGlyValPheLysGlyProAsp 1900
      :::::|||||
QY 1903 -----CCGAAAACCGCTTTGATACACCGTCGATTTTGAACCTTT 1944
      :::::|||||
Db 1901 GlyPheGluTyrPheAlaProAlaAsnThrGlnAsnAsnAsnIleGluGlyGluAlaIle 1920
      :::::|||||
QY 1945 GTA-----CGTGAACATCTCTTTCAGCGT----- 1968
      :::::|||||
Db 1921 ValTyrGlnSerLysPheLeuThrLeuAsnGlyLysLysTyrTyrPheAspAsnAsnSer 1940
      :::::|||||
QY 1969 -----GCCATTTTGAACAAATAAACATATCTATATAGACAGAGCT 2010
      :::::|||||
Db 1941 LysAlaValThrGlyTyrPargIleIleAsnAsnGlnLysTyrTyrPheAsnProAsnAsn 1960
      :::::|||||
QY 2011 ACAATATGCATATATCCCTTAAACCTGAATTTGATGTATGTTATGATGATGATAT 2070
      :::::|||||
Db 1961 AlaIleAlaIleAlaVal-----GlyLeuGlnValIleAspAsnAsnLysTyr-----Tyr 1976
      :::::|||||
QY 2071 TTTACTGATGAA-----CTTAAATGAAACGTA 2100
      :::::|||||
Db 1977 PheAsnProAspThrAlaIleIleSerLysLysTyrGlnThrValAsnGlySerArgTyr 1996
      :::::|||||
QY 2101 TTTATATTAAGAAGCGCTAGAGAAATTTGCGACCATAGCTATTTGATGATGATGAT 2154
      :::::|||||
Db 1997 TyrPheAspThrAspThrAlaIleAlaPheAsnGlyTyrLysThrIleAspGlyLysHis 2016
      :::::|||||
QY 2155 CTTTATATGGAAGTGC-----ATCGGTTCAATTTTCACTTAACGACTATTT 2202
      :::::|||||

```

```

Db 2017 PheTyrPheAspSerAspCysValValLysIleGlyValPheSerThrSerAsnGlyPhe 2036
      :::::|||||
RESULT 7
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falicliparum Transmission-Blocking Target Antigen, Pf5230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-323-170B-2
Alignment Scores:
Pred. No.: 1 39e-08 Length: 3135
Score: 183.00 Matches: 155
Percent Similarity: 33.22% Conservative: 134
Best Local Similarity: 17.82% Mismatches: 304
Query Match: 4.50% Indels: 277
DB: 1 Gaps: 38
US-09-727-892a-2 (1-2286) x US-08-323-170B-2 (1-3135)
QY 187 AGTTGCAATCTTTTATGAGCATTTTATACGATGTGAAAAGACGTGAT----- 237
      ||| ||| |||||
Db 631 SerValGluLysLeuTyrAspAsnIle---GluTyrValProLysLysSerProTyrVal 649
      :::::|||||
QY 238 ACAATCAAAATCAAAAACGATAT----- 264
      :::::|||||
Db 650 ValLeuThrLysGluGlnThrLysLeuLysGluLysLeuLeuSerLysLeuIleTyrGly 669
      :::::|||||
QY 265 ATCATGATGACATTAACGTATATTAATACGATATCATTT----- 306
      :::::|||||
Db 670 LeuLeuIleSerProThrValAsnGluLysGluAsnAsnPheLysGluGlyValIleGlu 689
      :::::|||||
QY 307 -----TACTTAAGACACCATGGCTATTT----- 333
      :::::|||||
Db 690 PheThrLeuProProValValHisLysAlaThrValPheTyrPheIleCysAspAsnSer 709
      :::::|||||
QY 333 ----- 333

```

Db 710 LysThrGluaspAspAsnLysGlyAsnArgGlyIleValGluValIleValGluPro 729
QY 334 -----GATATATTTACCGCGAAATATATAT 360
Db 730 TyrGlyAsnLysIleAsnGlyCysAlaPheLeuAspGluAspGluGluLysTyr 749
QY 361 TTAATCTCGAGAGAAATGACACATTAATAATGAAGAGCTACTATTAGCC 420
Db 750 GlyAsnGlnIleGluGluAspGluHisAsnGluLysIleLysMetLysThrPhePheThr 769
QY 421 AAAAT----- 426
Db 770 GluAsnIleTyrLysLysAsnAsnIleTyrProCysTyrMetLysLeuTyrSerGlyAsp 789
QY 427 ---CAAATGTAATTTAGAAAAACGTGTTAAATCTTCATCAATTTAGATTACAAATG 483
Db 790 IleGlyGlyIleLeuPheProLysAsnIleLysSerThr-----ThrCys 804
QY 484 TTTTAATGGTTTAAATTTATATGATACCTTTGAACCAATACATCAATT 543
Db 805 PheGluMetIleProTyrAsnLysGluIleLysTrpAsnLysGluAsnLysSerLeu 824
QY 544 GCAACATTA----- 552
Db 825 GlyAsnLeuValAsnAsnSerValIleTyrAsnLysGluMetAsnAlaLysTyrPheAsn 844
QY 552 ----- 552
Db 845 ValGlnTyrValHisIleProThrSerTyrLysAspThrLeuAsnLeuPheCysSerIle 864
QY 553 -----GGTAGAATTAATCTGATGCGTGTATTTAAGATCAGATCAGACTT 597
Db 865 IleLeuLysGluGluGluGluSerAsnLeuIleSerThrSerTyrLeuValIleValSerIle 884
QY 598 AAAACAGATTTATATACGATTTTGAT----- 627
Db 885 AsnGluGluLeuAsnProSerLeuPheAspPheTyrGluSerPheValProIleLysLys 904
QY 628 -----AAAGTATATGATATGATAGATAGAGCCATAGCATGCTGTCAGAA 675
Db 905 ThrIleGlnValAlaGlnLysAsnValAsnAsnLysGlu---HisSpirThrCysAsp 923
QY 676 TGTTCGAAACTCACACCTGACCACTTACATTCATATAGCAGTATATATTA 735
Db 924 PheThrAspLysLeuAspLysThrValProSerThrAlaAsnGlyLysLysLeuPheIle 943
QY 736 GGATAGCCCATTTTCATATATAGTATATATCCAAATTTGACATATAACAATTAGA 795
Db 944 CysArgLysHisLeuLysGluPheAspThrPhe-----Thr 955
QY 796 TTTTCATTAATATATGAAATCT---TACTGGAATATGAAATGACAGCTTTTCAGTTA 852
Db 956 LeuLysCysAsnValAsnLysThrGlnTyrProAsnIleGluIlePheProLysThrLeu 975
QY 853 CTACACCAATATCAAGATATTAATAATCTTATACACATTATCATTTCCATGATAGAT 912
Db 976 LysAspLysLysGluValLeuLysLeu-----AspLeuAsp 987
QY 913 TTT---TATGACTATATTAATATCATCTATCGTGGTGTAAATATGATATGATACCAATA 969
Db 988 IleGlnTyrClnMetPheSerLysPhePheLys-----PheAsnThrGlnAsnAlaLys 1005
QY 970 TACATTAACAACAACTAATGATGAGCCCTGTTTCTATTGACATCAATTCGAGTTATCCT 1029
Db 1006 TyrLeuAsn-----LeuTyrPro 1011
QY 1030 TATGTATGATCATGAAAAAATTCACATGTTATATCTTTAGCAACATATTCAGAA 1089
Db 1012 TyrTyrLeu-----IlePheProPheAsnHisIleGlyLys 1023
QY 1090 CCAACGTTA---ATCCCTACTTTTTCAGAT-----GATGACAAATATTTTTCATTA 1137

Db 1024 LysGluLeuLysAsnAsnProThrTyrLysAsnHisLysAspValLysTyrPheGluGln 1043
QY 1138 TATAAGATGATAAAGATGATATTTAAGCATGATATTAATTAATTAAT----- 1185
Db 1044 SerSerValLeuSerProLeuSerSerAlaAspSerLeuGlyLysLeuAsnPheLeu 1063
QY 1186 -----AANTCAGGTGATTTAGCTCAATGATTTGAAATACTATTAATATGATATGAT 1239
Db 1064 AspThrGlnGluThrValCysLeuThrGluLysIleArgTyrLeuAsnLeuSerIleAsn 1083
QY 1240 TACGTTAATATCAATCAATATCATTTAGAAATGATTCAGACATTAACGGTATTCATTCG 1299
Db 1084 GluLeuGlySerAspAsnAsnThrPheSerValThrPheGlnValPro-----ProTyr 1101
QY 1300 ATGCATATACGTTAAATTCGTTGTTATATATGATGATGATGATGATGATGATGAT 1359
Db 1102 IleAspIleLysGluProPheTyrPheMetPheGlyCysAsnAsnAsnLysGlyGluGly 1121
QY 1360 ATT-----ATTTCCAAACATATTTTATTAACACAGGTAAGTATAAAACAAATC 1413
Db 1122 AsnIleGlyIleValGluLeuLeuIleSerLysGlnGluLysIleLysGlyCysAsn 1141
QY 1414 AATATGACATCACCTTAGCATAT-----CACATTAAGATGATATCAGCA--- 1461
Db 1142 PheHisGluSerLysLeuAspPyrPheAsnGluAsnIleSerSerAspThrHisGluCys 1161
QY 1462 -----CACCATCTCAAAATGAGAGGTTATGTTATCTAAAGTCGTTTAAATGATTA 1515
Db 1162 ThrLeuHisAlaTyrGluAsnAsp-----Ile 1170
QY 1516 TATGCATACCTGCAATTCAGT---TCACATTTTACTTATTCGTTTATGAT---GATTAAC 1569
Db 1171 IleGlyPheAsnGlyLeuGluThrThrHisProAsnGluValGluValGluValGluLys 1190
QY 1570 AATGACATATAC-----AATATCATTAACGGTTACAAACACATGCAAGTAAAT 1617
Db 1191 AlaGluIleThrIleLeuLysAsnAlaGlnThrTyrAsnIleAsnAsnLysLysThr 1210
QY 1618 ATATATCTCTACATTTGTCACATCAGCTTCATGTTACTTA----- 1662
Db 1211 ValAspIleThrIleLeuLysAsnAlaGlnThrTyrAsnIleAsnAsnLysLysThr 1230
QY 1663 -----TTGGTTCCTTTCCAAATACTTAACGAAAGTGAATGACAGCAAT 1707
Db 1231 ProThrPheLeuLysIleProProTyrAsnLeuLeuGluAspValGluIleSer----- 1248
QY 1708 TTTATTTATTCGATACGATAGTATGATATGAAATCCGTTGTTAAACCTTATTCGAT 1767
Db 1249 -----CysGluCysThr-----IleLysGlnValValLysLysIleLysVal 1262
QY 1768 CCCAGTTTATTCGACCCGATAGCCTTAGTAAATGAGATATTTGAAGACAGACATA--- 1824
Db 1263 IleIleThrLysAsnAspThrValLeuLeuLysArgGluValGlnSerGluSerThrLeu 1282
QY 1825 ---GATAAGATGTTGTACTGATCATATAGAAGAAATATGATGAAAGATGAAAGATTT 1861
Db 1283 AspAspLysIleTyrLysCysGluHisGluAsnPhe-----IleAspProArgVal 1299
QY 1882 AAAATTCCTCTGCTGATACCCGAAACCCCTTTGATACAGCGTCGATTTGGAAC 1941
Db 1300 -----AsnLysThrPheAspGluAsnValGluTyrThrCys 1311
QY 1942 TTTGATGCT---GAACATTCCTTTGACGTCGATATATGAAACAAATAAAGATATGAT 1988
Db 1312 AsnIleLysIleGluAsnPhePhe-----AsnTyrIleGlnIlePhe 1325
QY 1999 AATGACAGGTATACATATCATATATCCGCT----- 2021
Db 1326 CysProAlaLysAspLeuGlyIleTyrLysAsnIleGlnMetTyrTyrAspIleValLys 1345
QY 2032 AAACTGAATGTATGATGATATGATATGATATTTATGATGATGATGATGATGATGAT 2091
Db 1346 ProThrArgValProGlnPheLysLysPheAsnAsnGluGluLeuHisLysLeuIlePro 1365

QY 913 TTT---TATGATATATTAAATCATTCATCGTGGTGTAAATATGTATACACCAAA 969
 Db 988 TlEglnIyrglnmetPheSerLysPhePheLys-----PheasnHrglnasnaIalys 1005
 QY 970 TACATAAACAAATTAATGATGAGCCCTGTTTCTTATGACATCAATTCGATTAACCT 1029
 Db 1006 TyrlEuaSn-----LeuTyrrPro 1011
 QY 1030 TATGTATGTATCATGAAAAAATCCACATGCTATACCTTTACGACACCTATTACGAA 1089
 Db 1012 TyrrTyrlEuu-----TlEphEproPheasnHslEgIlyLys 1023
 QY 1090 CCAACGTTA-----ATCCCTACTTTTTCATGAT-----GATGACATTAATTTTCATTA 1137
 Db 1024 LysglIueuLysasnsnProHrTyrlLysasnHslLysaspyalLysTyrrPhegluGln 1043
 QY 1138 TATAGATGTATAAGATGATTAACGATGATTTTATTAATAAATY----- 1185
 Db 1044 SerSerValLeuSerProLeuSerSerAlaAspSerLeuLysLeuLeuAsnPhelEu 1063
 QY 1186 -----AATCAGGTGATTAAGCTCAATGATGTAAATGATAATGATTAATGAT 1239
 Db 1064 AspHrglnGlnIuHrValCysLeuHrglnLysIleHrTyrlEuaSnLeuSerIleasn 1083
 QY 1240 TACGTATATATCAATACAAATACATTAAGAATGATTCAGACATTAAGGATATTGATGC 1299
 Db 1084 GlueuLysSerAspAsnAsnHrPheSerValHrPhegluValPro-----ProTyrr 1101
 QY 1300 ATGATTAATAGCTGTATATGCTTTGTATATGAAATGAAATGAAATCTTACATGACCTGAT 1359
 Db 1102 TlEaspIleLyscluprPheTyrrPheMetPheglIcysasnAsnAsnLysglIyglIy 1121
 QY 1360 ATT-----ATTTTCAAAACCTTTTATTAATAACAAGGTAAAGTAAACAAATAC 1413
 Db 1122 AsnIleglIyIleValGlnLeuLeuIleSerLysglIuGlnLysIleLysglIyCysasn 1141
 QY 1414 AATATGATCACCCTACCTACCTAT-----CACATTACGATATATATCAACGAA--- 1461
 Db 1142 PheHslcIueSerLysLeuAspTyrrPheasnGlnAsnIleSerSerAspHrHsgLycys 1161
 QY 1462 -----CACCCATCACTCAATGAGAGGATATGTATCTAAAGTCCCTTAAATGATTA 1515
 Db 1162 ThrLeuHslalEtyrglnLysnAsp-----Ile 1170
 QY 1516 TATGGCATACCTGCATTAACCT---TACATTTTAACTTATTCGCTTAAT---GATAC 1569
 Db 1171 TlEglYpheaSnCysLeuGlnHrHrHsPheAsnGlnValGlnValGlnValGlnAsp 1190
 QY 1570 AATGACTATATC-----AATATCATTAACGCTTACAAACAAACACTGACCTAAT 1617
 Db 1191 AlagIuIeTyrlEuaGlnProGlnAsnCysPheasnAsnValTyrlLysglIeuaSnSer 1210
 QY 1618 ATATTTATCTCTACATTTGTCACATGACGTTCATGTATTAACCTTA----- 1662
 Db 1211 ValAspIleHrHrHrIleLeuLysAsnAlaGlnHrTyrlLysasnLysnLysTyrr 1230
 QY 1663 -----TTGGTTCCTTCCAACTACTTAACGGAAGTAAATGACGCAAT 1707
 Db 1231 ProHrPheLeuLysIleProProTyrrAsnLeuLeuGlnAspValGlnIleSer----- 1248
 QY 1708 TTTATTTATTCGATACGTAGTATGTTATATGAAATCCGCTGTTTAAACCTTATGAAAC 1767
 Db 1249 -----CysglncysThr-----TlEysglValValLysLysLysLysVal 1262
 QY 1768 CCCAGTTTATGACCCGATAGCCTTAGGTAATGAGATATTGAAAACACAGATGAT--- 1824
 Db 1263 TleIleThrLysasnAspHrValLeuLysArgGlnValGlnSerGlnSerHrLeu 1282
 QY 1825 ---GATAAGATGTTTGTACTGAAATCAATAGAAATATGATGAGTGAATGGAAGAT 1881
 Db 1283 AspAspLysIleTyrrLysCysGlnHsgLysnAsnPe-----TlEasnProHrVal 1299
 QY 1882 AAAATGCTTTCCTGCTGTATACGAAAAAGCCCTTGTGATACAGCGCTGATTTTGAAC 1941

Db 1300 -----AsnLysThrPheasnPyGlnAsnValGlnTyrrHrCys 1311
 QY 1942 TTTGACGT---GAACAATTCCTTGACGCGTGCCATTATGAAACCAATAAAGATCTAT 1998
 Db 1312 AsnIleLysIleGlnAsnPhEpe-----AsnTyrlEglnIlePhe 1325
 QY 1999 AATGAGCAAGGTACATATCATATCCGCT----- 2031
 Db 1326 CysProAlaLysAspLeuGlnTyrlEtyrrLysasnIleGlnMetTyrrAspIleValLys 1345
 QY 2032 AAAAGTGAATGTATGTCGTATGATATGATATGATATGATTTTACTGATGACCTAATATG 2091
 Db 1346 ProHrHrArgValProGlnPheLysLysPheasn-AsnGlnGlnLeuHslLysIleLepr 1365
 QY 2092 AAACGTGAATTTATATTTAAAGACGCTAGACAAAATTCGACCATGATCAATTTGATGAT 2151
 Db 1365 AsnSerGlnLeuLysHslLysTyrrLysgluMet-----LeuIleLe 1379
 QY 2152 ATCTTTTATGTAAGATGACATCGTTCATCTTTCATTAACGACTTATTTCCAGTTGAA 2211
 Db 1379 TyrrAsnGlnGlnLysValAspLeuLeuHslPheTyrrValPheLeuProIleTyrrIle 1399
 QY 2212 CGTTCAGTACATTAACAATCTGATTTGC 2239
 Db 1399 sasplEtyrglnupheAsnIleValCys 1408
 RESULT 9
 US-09-134-001C-4352
 ; Sequence 4352, Application US/09134001C
 ; Patent No. 6580370
 ; GENERAL INFORMATION:
 ; APPLICATION: Lymn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4352
 ; LENGTH: 944
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4352
 Alignment Scores:
 Pred. No.: 1.29e-08 Length: 944
 Score: 181.50 Matches: 155
 Percent Similarity: 38.61% Conservative: 128
 Best Local Similarity: 21.15% Mismatches: 226
 Query Match: 4.46% Indels: 224
 DB: 4 Gaps: 37
 US-09-727-892a-2 (1-2286) x US-09-134-001C-4352 (1-944)
 QY 497 TTAATTTAATATATGATTAATCACTTTTGAAGAAACCAATACATCAATGCAATTAAGTA 556
 Db 286 LeuMetLeuMetTyrrLeuLeuSerLeu---LysCysLeuTyrrLeuLeuSerHs---Ser 303
 QY 557 AGAATTTACTGANGTGCGTATTTTACAGAGAAATCACAACCTTAAACAGATT-----TTA 610
 Db 304 ThrMetTyrrLeuLeuPheProSer---ArgSerHslLeuLeuMetGlnMetHslTyrrArg 322
 QY 611 ATTATACATTTTGTATGAATAGATATGATGATGATGATGAGAGCCTATGACTATGCTG 670
 Db 323 MetLeuMetTyrrLeuLeuSerLeuLysArgLeuTyrrLeuMetSerHslMetCys 342
 QY 671 TGAATGTTTTCGCAAAACTCACACCTGTAACACTTATACATTCATTAATGACGTGATTA 730

Db 342 s---TrpHisCysLeuSerHisSer-HisSerLeuMetLeuMetHis-----CysL 358
 QY 731 TATTAGTATGTCACATATTCATTATAGTAT----- 762
 Db 358 eumSerMetLysSerLeuAsnLeuLysAspSerLysLeuSerSerHisLeuArgMet 378
 QY 763 -----ATATTCCAAATTTTGACTATACAAATTACA----- 795
 Db 378 yLeuLeuPheProSerArgSerHis--LeuLeuThrLeuMetHisCysLeuMetSer 397
 QY 796 --TTTTCATGGAATATATGGAATCTTACTGTAATATGAATGACACGTTT----- 846
 Db 397 eLysSerLeuAsnLeuLysAspSerLysLeuSerSerHisLeuThrMetLysLeuHis 417
 QY 847 -----CAGTTACTCA 856
 Db 417 yLeuThrHisSerHisSerLeuMetThrPheHisCysArgMetSerThrArgLeuLeu 437
 QY 857 ACCAATATCAAGATATATAA-----ATATCTTATACACATT 892
 Db 437 sn-----LeuLysArgLeuCysLeuSerThrArgLeuMetThrPheHisL 453
 QY 893 AT-----CATTCATGATGATGATTTTATGACATATATTAATCATTCATCGGTG 946
 Db 453 eLeuSerHisSerHisSerLeuThrLeuMetHisCysArgMetSerThrTyr-----S 471
 QY 947 GTTTATATGATATACACCAATATACATACAAACTAATGATGACGCTGTTTCTA 1006
 Db 471 eLeuAsnLeuLysArgSerLysLeuMetLeuHisLeu--TrpMetLysLeuHisPhePro 490
 QY 1007 TTGACATCAATTCAGATTCCTATGTCATGATGATGATGATGATGATGATGAT 1066
 Db 491 SerArgSerHisSerLeuMetLeuMetHisCysArgMetSer----- 504
 QY 1067 ACTTTTCGAAACACTATTCGAAACCAACGTTATCCCTACTTTTATGATGATGACATT 1126
 Db 505 -----ThrArgLeuLeuAsnLeuLysArgLeuCysLeuSerThrArgLeuMetTh-- 521
 QY 1127 ATTTTCATATATAGATGATGATA-----AAGATGATTTTAAAGATGATTTTATATA 1180
 Db 522 -----TyrPheHisLeuLeuSerHisSerHisSerLeuThrLeuMetHisPheLeu 538
 QY 1181 AAATTAATGACCGGTATTCAGTCAATGATGTAATGATGTAATGATGATGATGATGAT 1234
 Db 539 ThrLeuThrLysLeuLeuSerProLysAspLeuCysLeuSerLeuHisLeuMetThrTyr 558
 QY 1235 -----ATGATTACGTTATATATCAATACAAATACATTTAGAA 1270
 Db 559 LeuThrCysLeuThrHisSerHisSerLeuThrLeuMetHisCysArgMet----- 575
 QY 1271 TGATTCAGACATTACGGGATTTGATTCATGCAATATACGTTGTTATTCG----- 1321
 Db 576 -----SerThrLysSerLeuAsnLeuLysCysLeuThrLeuMetLeuHisLeuThrMet 593
 QY 1322 -----TTGTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369
 Db 594 CysThrPheHisCysLeuThrHisSerHisSerLeuMetThrPheHisPheLeuThr 613
 QY 1370 AAAACGATTTTATATAA-----CACACAG 1393
 Db 614 ArgLeuLeuAsnLeuLysArgLeuCysLeuMetLeuHisSerThrMetLysPheHisLeu 633
 QY 1394 GTAAGTTAAAAAACAATATCATATGACATCCTTACGATTCATGACATTCAGATGATA 1453
 Db 634 LeuSer-----HisSerArgLeuLeuThrLeuMetHisCys 645
 QY 1454 TCAACGACACCATCTACTCAATGAGAGGATTTATGATTAAGTGGTTTAAATGAT 1513
 Db 646 ArgMetSerThrHisSerLeuSerLeuLysAspSerLysLeuSerLeuHisLeuMetThr 665
 QY 1514 TATATGCAATCCGTCATTCAGTTCACATTTTACT----- 1549
 Db 666 TyrLeuThrCysLeuThrHisSerHisSerLeuThrLeuMetHisCysArgMetSerThr 685

QY 1550 -----TATTCGCTTTAGATGATACAAATGACATGATA 1582
 Db 686 HisSerLeuSerLeuLysAspSerLysLeuSerLeuHisLeuMetThrLysPheThrCys 705
 QY 1583 ATA-----TCATTAACGTT-----ACAAATACAGTGAAC 1612
 Db 706 LeuLeuHisSerArgSerLeuThrLeuMetHisCysArgMetSerThrLysSerLeuAsn 725
 QY 1613 GTA-----ATATTTATCTCTACATTTGTCACATTCAGCTTCATTTATTAAT----- 1660
 Db 726 LeuLysArgLeuThrLeuSerLeuHisLeuMetThrLysLeuHisCysLeuThrHisSer 745
 QY 1660 ----- 1660
 Db 746 HisSerLeuMetThrPheHisCysArgMetSerMetLysLeuLeuAsnLeuLysArgLeu 765
 QY 1661 -----TATTCG-----TTCCCTTCACATTC 1681
 Db 766 CysLeuSerLysLeuLeuMetThrLysLeuArgLysThrPheSerHisLeuArgSerLeuThr 785
 QY 1682 TAACGGAAGTGAATTCAGCACA-----ATTTATTTATTCGCAATGATGATGAT 1732
 Db 786 LeuMetArgLysThrPheSerThrLysSerLeuLeuLeuLeuLeuLysLeuLysLeu 805
 QY 1733 TGATATGAAATCCGCTGTTAAACCTTTATGACCCGAGTTTATTCGACCCGATAGCCT 1792
 Db 806 LeuLeuArgSerLeuLeuLeuMetArgLysThrPheSerMetCysSerLeuSer----- 823
 QY 1793 TAGGTAAATGGCATTTGAAACGAAACAGATAGATAGATGTTTGTACTGAAATCATAGA 1852
 Db 824 -----LeuLysCysLeuThr-----LeuLeuSer 831
 QY 1853 AATATGCATATGAGATGATGGAAGATTAATAATTCCTGCTGATACCGGAAAAACG 1912
 Db 832 HisSerHis-----LeuLeuMetLeuThrHisPheLeuThr 843
 QY 1913 CCTTTGATACAGCGGTGATTTGAAACCTTTGATGCGTGAACAAATCTTTGACGGTGCA 1972
 Db 844 ---LeuThrArgLeuLeuAsnLeuLysArgLeuCysLeuSerLeuHisLeuMetThrTyr 862
 QY 1973 TTATTTAAAAACAATAAAGATCTATATGAGCAAGGATGATGATGATGATGATGATGAT 2032
 Db 863 PheThrSerLeuThrHisSerPheLeuLeuThrLeuMetHisLysGlnMetSerThrTyr 882
 QY 2033 AAACGTAATTTGATGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2092
 Db 883 SerLeuLeuLeuLeuLeuLysLeuCysLeuLeuLeuLeuArgSerLeuLeu----- 900
 QY 2093 AACGTAATTTATTAAGAACGCTAGAGAAATTTGCAACATAGTCAATTTGATGATA 2152
 Db 901 -----ThrLysThrSerThrSerThrHisLeuThrThrLeuLysAspSerLysLeu 917
 QY 2153 TTCTTATATGGAAGTGAATGTCATGTTTACCTTAACGATTAATTTCCAGTTGAAC 2212
 Db 918 LeuLeuMetLeuLeuAsp---SerLeuLeuHisLysHisLeuLeuLeuLeuLysLeuLeu 935
 QY 2213 GTTCAGTACTACAAAT-----CTGATTTGC 2239
 Db 936 -----TyrLeuLeuLysHisLysLeuLeuCys 944
 RESULT 10
 US-09-134-001C-4452
 ; Sequence 4452, Application US/09134001C
 ; Patent No. 6380378
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 4452
 LENGTH: 956
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4452

Alignment Scores:
 Pred. No.: 2,456-08
 Score: 178.50
 Percent Similarity: 33.57%
 Best Local Similarity: 19.71%
 Query Match: 4.39%
 Gaps: 41

US-09-727-892a-2 (1-2286) x US-09-134-001C-4452 (1-956)
 QY 286 AATPAATGATATATATTTTACCTTAAGACACCATGTTATTTGATATATATTA 345
 Db 191 AsnGluTyrLysSerPheGluTyrGlnThrLeuGlnLysValPheAspSerThrVal 210
 QY 346 CGGAAATATATATTTAAATCTGCAGAAAGAAAT 381
 Db 211 ValGlnAsnSerGluLeuLysPheAsnGluSerLysLeuLysLysProAsnLeuMet 230
 QY 382 --GAACACACATTAATAATGAGAG 405
 Db 231 GlnGlnHisAlaLeuLysSerLeuGlnSerLeuArgAsnValGlyGlyLysGlyLeu 250
 QY 406 -----GCTACTATTTTAGCCAAATCAAAATGCTAATTTTGAAGAAAGCTGTTAA 456
 Db 251 IleIleSerAlaThrGlyThrGlyLysThrIleLeuCysAlaLeuAspAlaArgAlaTyr 270
 QY 457 TCTTCATCATTTAGATTTTACAAATGTTTAAATGCTTTAAATTAATTAATTTGAT 516
 Db 271 SerProAspLysPheLeuPheIleValHisAsnGluGlyIleLeuAsnArgAlaIleGlu 290
 QY 517 AACTTTATGAAA 555
 Db 291 GluPheLysLysValPheProTyrGluAspGluSerAspPheGlyLeuLeuThrGlyLys 310
 QY 556 AAGAATTAAGTGGTGGTTATTTA 609
 Db 311 ArgLysAspHisAspAlaLysPheLeuPheAlaThrIleGlnThrLeuSerLysGlu 330
 QY 610 AATTAACGATTTTGATTAAGTAATGATGATGATGATGATGATGATGATGATGAT 669
 Db 331 AsnTyrLysPhePhe-----AsnSerAsnHisPheAspTyrIle 343
 QY 670 GTG-----AATGTTTGCAGAAAGCT 690
 Db 344 ValPheAspGlnAlaHisArgThrAlaAlaSerSerTyrGlnLysIlePheAsnTyrPhe 363
 QY 691 ACACCTGAACCACTACATCATCATTAATGACGATTAATTTAGTATGTC----- 744
 Db 364 LysProAsn-----PheLeuLeuGlyMetThrAlaThr 374
 QY 745 -----CATATTCATTTATGATATTTTCCAAATTTTGCATTAACAAATTAACATTT 798
 Db 375 ProGlnArgThrAspGluLeuAsnIlePheGluLeuPheAsnTyrAsn-----IleAlaTyr 393
 QY 799 TCATGATATATGTAATCT----- 819
 Db 394 GlnIleArgLeuGlnGlnAlaLeuGlnSerAsnIleLeuGlyCysProPheHisTyrPheGly 413
 QY 820 -----TACTGATATATGAAATGACAGCT----- 843
 Db 414 ValThrAspTyrIleGlnHisGlnIleMetSerGlnGluAspAlaPheAsnLeuLysTyrLeu 433
 QY 843 ----- 843

Db 434 AlaSerAsnGluArgValGlnHisIleIleLysLysThrAsnTyrTyrGlyTyrSerGly 453
 QY 844 -----TTTCAGTACTCAAC 858
 Db 454 AspValValLysGlyLeuIlePheValSerSerArgGlyAlaTyrGlnLeuAlaAsn 473
 QY 859 CATATCAAGATATTAATAATCTTATACATTTATTCATTTCCATGATGATGATGATGAT 918
 Db 474 GlnLeuSerLysArgGlyIleSerSerValGlyLeuThrGlyLysAspSerIleAlaTyr 493
 QY 919 -----GACTATATTTAAATCATCTATCGTGGTGTAAAT-----ATGATTAACACCA 969
 Db 494 ArgThrGlnThrIleGlnGlnLeuLysGluGlySerIleAsnTyrIleIleThrValAsp 513
 QY 970 TACATAAACAACATTAATGAGAGCGCTTTTCTATGACATCAAT----- 1017
 Db 514 LeuPheAsnGluGlyIleAspIlePro-----GlnIleAsnGlnValValMet 529
 QY 1018 -----TCAGATTATCCTTATGATGATGAT----- 1041
 Db 530 LeuArgProThrLysSerSerIleIlePheIleGlnGlnLeuGlyArgGlyLeuArgLys 549
 QY 1042 -----CATGAAAAATTCACACATGCTTATCTTTCAGACACATTCGAAACCAAC 1095
 Db 550 SerThrAsnLysGluPheValThrValIleAspPheIleGlyAsnTyrLysThrAsnTyr 569
 QY 1096 TTAATCCCTACTTTT-----GATGATGACAAATTTATTTTCAATTA 1137
 Db 570 MetIleProIleAlaLeuSerGlyAsnLysSerGlnAsnLysAspAsnTyr-----ArgLys 588
 QY 1138 TATAGATGTTAAAGATGTTTACGATGATTTATTAATAAAT----- 1185
 Db 589 PheLeuThrAspThrThrValLeuAsnGlyValSerThrIleAsnPheGluValAla 608
 QY 1186 AATACGCTGATTAACG-----CAATGATTTGA 1215
 Db 609 LysAsnLysIleTyrAsnSerLeuAspSerValLysLeuAsnGlnProLysLeuIleLys 628
 QY 1216 AATATCTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1275
 Db 629 GlnAlaPheAsnAsnValLysAspArgIle-----GlyLysLeuProLeuLeu 644
 QY 1276 CAAGCATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
 Db 645 MetLysPheIleAsnAsnAsp-----SerIleAspProSerValIlePheSer 660
 QY 1336 TGTGAATACTTTCATGACGATATTTTTCGAAACATTTTATTAACAACAAGGT 1395
 Db 661 -----ArgPheLysAsnTyrTyr-----GluPhe 668
 QY 1396 AAGTTAAACAAATCAATATGATACATCACCTTACGACATCATCTACTGATGATATC 1455
 Db 669 LeuIleLysAsnLysIle-----IleGluAsnGlnLeu 679
 QY 1456 AACGACACACCTACCTCAATGAGAGGATTAATCTTAAGTCGTTTAAATGATTA 1515
 Db 680 SerIleAsnGlnLysPheAsn-----LeuThrPheLeuSerArgGlnLeuSerProGlyLeu 698
 QY 1516 -----TATGGCATACCTGCTTACGT-----TCACATTTTAAAC 1548
 Db 699 LysLysValAspIleAspValLeuLysGlnValIleGlnAsnAspValThrTyrGluAsn 718
 QY 1549 TTATTCGCT-----TAGATGATTAACATGACATA-----TACATATTCATTAACGCT 1596
 Db 719 LeuThrLysLysMetLeuAsnIleAsnAspIleSerGlnTyrAspIle----- 735
 QY 1597 TACAAAAACACTGAAGCTAATATATTTATTCCTACATTTGTC-----ACATCAGCTTCA 1650
 Db 736 -----AsnThrSerLeuSerIleLeuAspPheThrPhePheLysLysThrIleGlyLys 753
 QY 1651 TTGATTAAGTATGTTGCTTCTTCCATTAAGGAAAGTGAATGACGACATTTT 1710

```

Db 754 ThrtgylleuproLeu-----ileglnTy-----Lysaspasleu 766
QY 1711 ATTATTCGATCTGATGACTGTTGTATATGAATCCGTTGTAACCCCTTATGAACCC 1770
Db 767 ilecysleuAlaasnlu-----PheLysgluAlaleuasnLysProleupheasThr 784
QY 1771 AGTTATTCGACCCGATAGCCTTAGTAATGGATTTGAAAAAGACAGATAGATAAG 1830
Db 785 PhevalHisaspLeuileaspLeuAlaasnTyAsnaspArgTygluAsnLysLys 804
QY 1831 ATGTTTCTACTGAATCATTAAGAAATATGATATGAACTGATGAGAAAGATTAATTTGCT 1890
Db 805 AsnserleuileuTyAsnLysTySerArg----- 815
QY 1891 TCCTCTGATATACCGAANAAGCGCTTGATACAGCGTCGATTTGAAACCTTTGTACGT 1950
Db 816 -----GluaspPhevalLys 820
QY 1951 GAACAAATCTTTGACGGTGCCATTATTGAANAACATAAAGTATCTATAATGAGCAAGT 2010
Db 821 LeuLysasnTyasp-----LysaspGluSerGly 830
QY 2011 ACATATTCGATATTCGCGTCTAAACGAAATTCATGCGTAATGATATGATGATAT 2070
Db 831 ThrleasnGlyTy----- 835
QY 2071 TTTACTGATGAACCTTAATATGAAACGTGA-----TTATATTAAGAACGCT 2118
Db 836 -----ArgmetLysHisArgThrleuProleupheleThrTyaspLys 850
QY 2119 AGACAAATTTTC---GACCATAGTCATTTGATGATATCTTTATGAAAGACATC 2175
Db 851 HisgluasnileSeraspasnThrLysTyaspGluPheleuSerGluaspGluLeu 870
QY 2176 GGTTCATTTCCACTTAAC-----GACTATATTCGAGTGAAGCT 2214
Db 871 LysThrTyThrArgSerHisArgLysleuThrSerProgluValGluAsnIleleuLys 890
QY 2215 TCAGTACATACAAATCTGATTTGCATATATTAAGCTGAACATGATGAA 2265
Db 891 HisgluGluAsnAsnThrAspmetTyrlIlePheValLysLysArgaspasp 907

RESULT 11
US-08-645-193B-15
; Sequence 15, Application US/08645193B
; Patent No. 5962253
; GENERAL INFORMATION:
; APPLICANT: Kupke, Thomas
; APPLICANT: Gotz, Friedrich
; APPLICANT: Kempler, Christoph
; APPLICANT: Jung, Gunther
; TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
; TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645,193B
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893

```

```

; REFERENCE/DOCKET NUMBER: 0652.1540000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-645-193B-15

Alignment Scores:
Pred. No.: 3,06e-08 Length: 990
Score: 177.50 Matches: 162
Percent Similarity: 35.20% Conservative: 146
Best Local Similarity: 18.51% Mismatches: 270
Query Match: 4,36% Indels: 297
DB: 2 Gaps: 45

US-09-727-892a-2 (1-2286) x US-08-645-193B-15 (1-990)
QY 298 AATCATTTTCTACTTAAGACACACGCGTTATTTGATATATTAATGACGCAAAATATA 357
Db 28 AsnGlnPheleuLysSeraspasnIleaspTy---AspleuIleleuGluAsnaspIle 46
QY 358 TATTTAAATCT----- 369
Db 47 PheLysgluSerIleMetThrThrThrTyAsnleuTyGlnSerIleGlyLysIleasp 66
QY 370 ---GCAGACAAATGACACACACATTAATAATGAAGCGCTACTT----- 414
Db 67 TrpGluLysaspasnLysLysThrArgasnValLysleuSerleuLysTyLysIle 86
QY 415 -----TTA 417
Db 87 ArgmetSerThrArgSerThrProTyGlyMetLeuSerGlyValAlaLeuGlyLuphe 106
QY 418 GCCAAATGCAAAATGTAATTTGAAGAAACGCTT-----AAATCTGAATCATTTA 471
Db 107 SerGluasnasnIleLysIleLysaspSerSerPheHisLysTyaspValLysIle 126
QY 472 GAT-----TTACAAATGTTTAAATGATGTTTAAATTAATTAATTAAT 513
Db 127 AspGlyGlnThrleuTyrlLysleuValHisTyrlleuGlnSeraspTyThrTyLys 146
QY 514 GATACCTTTATGAAGAACCAATACATGCAATTCAGACATTTAGGTAAGAAATTAATGATGT 573
Db 147 AsperPheValIleThrArgasnGln----- 155
QY 574 GGTATTTAAGCAGATCACTTAACAGATTTTAATATATGATTTTGAATAAGAT 633
Db 156 AsnTyrlIleTyAsnAsnArgleuTyrlleuaspasnSerSerIle-----Thglu 173
QY 634 AATGATATGATGAT-----AGTGAAGCCTTAGCTAGCTGGAATGTTTGCAAA 687
Db 174 AsnLysArgasnAspValleuSerValLysTyAsnSerIleleuValPheIleHisglu 193
QY 688 CTCACACTGACCAACTTACTAC----- 711
Db 194 AsnserLysasnIleThrTyrlGluGluLeuValGlnleuIleSerSerLysTySer 213
QY 712 -----ATTCATAATGACGTGAT 729
Db 214 ilegluasnLysgluGluValLysValPheValGlnGluLeuIleasnLysGluIleIle 233
QY 730 ATATTAGTATGTCGCAATTCATATAGTATATATTTCCAAATTTGAC---TATAAC 786
Db 234 -----PheSerLysleuArgProThrleuGluAsnLysasn 245
QY 787 AAATTAACATTTTCATGATATATGGAATCT----- 819

```

```

Db 246 ProLeuaspTyrIleIleAsnSerLeuAsnProLysAsnSerLeuValGlyThrLeuIle 265
QY 820 TACTGATATGAATGACACGTTTCAGTTACTC-----AACCA 861
Db 266 AsnIleSerAsnGluIleThrLysTyrSerLysMetProLeuGlyLysGlyIuTyrLys 285
QY 862 TATCAAGATAT-----AAATATCTTATACATTCATTCATTCATCATGAT 906
Db 286 TyrLeuAspIleValAsnLeuMetSerGlnLeuPheValSerLysAsnTyrLeuGlnIle 305
QY 907 ATGATTTTATGACTATTAATATCATTTATCGTGGTGTTA-----AATATGAT 960
Db 306 AspThrTyrIleAspTyrSerArgAsnGlnLeuLysGlnSerLeuAlaAspAsnIleSer 325
QY 961 AACACAAATACATTAACAACATAATGATGAGCGCTGTTTCTATTCATTCACATTCATG 1020
Db 326 GluAlaIleTyrIleLeuThrPheLeuSerProAsnHisPheGlyThrIleArg 345
QY 1021 AGTTATCCTTATGATGATGATCATGAAAAAATCCACAGCTTATACCTTTTACGAAC 1080
Db 346 Asn-----TyrHisGlu-----PhePheMetAspLys 354
QY 1081 TATTCACACCAACGTTATC-----CCTACTTTTTCATGATGACATTCATTTTTC 1134
Db 355 TyrGlyPheGluGlnLeuValAsnLeuLysGlnLeuLeuSerAspIleAsnGlyPheGly 374
QY 1135 TTAATATAGATGATTAAGATGATTAATACGATTTTATTAATAAATTAATACCGT 1194
Db 375 TyrProLysLysAspSerTyrSerPheSerAsnHisIleAlaPhe---LeuLysGluLys 393
QY 1195 GTATTACGTAATGATTTGTAATAATCTATTAATATGATATACCTTATATCAT 1254
Db 394 TyrLeuLeuAlaIle-----GlnAsnAsnSerHisIleGluIleThr 407
QY 1255 ACAATATCATTAAGATGATTCACAGCATTCACGGTATGATTCATGATATACGTT 1314
Db 408 GluAsnAspValLysAsnLeuGlnLysAsnAsnThrValSerLysIleAsnAlaProVal 427
QY 1315 AATTCGTTTATATATGATGATGATAC-----1344
Db 428 SerThrGlnIleTyrSerGlnIleTyrPheGlyAsnSerIleLysGlyTyrGluAspPhe 447
QY 1345 -----TTTCATCGACGCGATGATTTTTC 1371
Db 448 AlaValIleSerProIleLeuGlySerPheAsnAlaGlyAlaThrPheGlyArgPheThr 467
QY 1372 ---AACTATTTTATTAACAACAAGTAAAGTTAAACAACAATC-----1413
Db 468 GlyAsnPheAsnIleLysLysLysAsnGlnLeuGlnLysGluIleValHisHisTyrAsn 487
QY 1414 AATATGACATGACCTTACGACTATCATCTACTGATGATATACAGAACACCCATAC 1473
Db 488 AsnTyrMetAsnGlnLysAsnLeuGlnIleSer---GlnLeuAsnGlnAlaProLeuAsn 506
QY 1474 AATGACGAGGTATGTTATCTAAAGTCGTTTAAATGATATATGCGATACCGCATTA 1533
Db 507 SerArgAsnValAsn-----IleLeuAsn-----514
QY 1534 CGTTACATTTTAACCTTATTCGTTTACGATGATTAACAATGACATATACATATC-----1587
Db 515 -----AsnAsnArgIleTyrAsnThrCysLeu 523
QY 1588 -----ATTACCGT-----1596
Db 524 AsnLeuAsnLeuProLysSerAspIleAsnAspIlePheIleGlyAlaThrPhe 543
QY 1597 -----TACAAAAACCTACAGCTAATATATATCTTACATTTGTCACATCAGCT 1647
Db 544 AsnLysLeuTyrLeuTyrSerGlnLysHisAspSerArgIleValPheVal---SerAsn 562
QY 1648 TCATTTGATATATTA-----TTGGTTCCCTTC-----1674
Db 563 SerMetPheAsnTyrGlnPheGlySerGlnLeuTyrLysPheLeuArgGlnIleSerPhe 582

```

```

QY 1675 -----CAATACTTAACGGAAGTGAATTTGACGACATTTTATTTAT 1716
Db 583 GluLysThrLysPheIleGlnProIleThrGlnGluGlyIle---AspSerLeuProPhe 601
QY 1717 TCGGATACGATGATGTTGATTAAGAAATCCGTTGTAACCCCTTA-----TTGAC 1767
Db 602 CysProArg---IleIleTyrLysAsnIleIleLeuLysProAlaThrTrpLysIleAsn 620
QY 1768 CCCAGTTTATTCGAC-----CCGATACCGCTTACGATAA 1800
Db 621 SerGlnMetPheSerGlnThrGlnAsnThrPheAsnArgPheAlaThrIleArgGluLys 640
QY 1801 TGGGATATGGAACCAACAGTA-----GATACATGTTTCTACTGAT---1845
Db 641 TrpHisIleProLysAspValIleIleAlaPheGlyAspAsnArgLeuLeuAsnLeu 660
QY 1846 -----CAT-----AAGAAATATGCAATATGAAAGTGAAGTGAATAA 1884
Db 661 LeuAsnAspLysHisLeuIleIleLeuLysLysGlnLeuLysLysHisGlyArgIleArg 680
QY 1885 ATTGCTTTCGCTGGTATACCGAAAAAGCGCTTGATPACAGCGTCGATTTT-----1935
Db 681 IleLeuGlnSerPheIleAsnGlnSerAsnGlnAlaArgMetLeuGlnIleValThrPro 700
QY 1936 -----GAAACCTTTGATACGTAACATTCCTTGACGGCGCATTAAT-----GAA 1980
Db 701 LeuTyrLysLysThrSerLeuLysGlnIleSerPhe-----IleIleProLysAsn 717
QY 1981 AACATTAAGATATGATATATGACGACAGCTACATATGATATAT-----2025
Db 718 ArgAsnLysHisPheAsnAsnLeuLysAspTrpPheSerIleHisLeuSerIleProLys 737
QY 2026 -----CCGCTTAACGGAATGATGAT 2049
Db 738 ThrTyrGlnAspAsnPheIleGlnAspTyrLeuLeuProPheIleThrGlnLeuVal 757
QY 2050 GGTAAATGATATGATGAATATTTACTGATGACACTTATATATGAACGGAATTTATA---2106
Db 758 AsnAsnPheIleAsnLysPhePheTyrIleLysPheLysGlnAspGluAspPheIleLys 777
QY 2107 TTAAGACGCTAGAGAAATTTGACCATAGTCAA-----TTT 2145
Db 778 LeuArgLeuLeuArgGlnAspLysPyrSerGlnIleTyrSerPheIleLysAsnTrp 797
QY 2146 GATGATATCTTATATATGAAGATGACATCGTTCAATTTTCACTTAACGACTTATTTCCA 2205
Db 798 LysAspTyrCysLeuLeuAsnSerGlnLeuTyrAspTyrSerIleValAspTyrValPro 817
QY 2206 GTTGACCTTCAGTACATACAAATCTGATTCATATATTA 2250
Db 818 -----GluValTyrArgTyrGlyGlyProHisValIleGlu 829

RESULT 12
US-09-091-117-5
; Sequence 5, Application US/09091117
; Patent No. 6171589
GENERAL INFORMATION:
APPLICANT: The University of Melbourne
TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
TITLE OF INVENTION: Vaccines
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

QY 1642 TCACGT-----TCATGATTAACCTATTGGT-----CCT 1671
Db 869 LysLysySGUAlaAsnSerSerProValLeuLysLeuIryrSapLaValIlySaYqAsnasp 888
QY 1672 TTCACATACTTAACGGAAGAATGGAATTGACAGACATTTTATTATATGCGCATGCTAGT 1731
Db 889 ArgIntYrYAcProAsnHisHisIleSAsPLeuAlaYqAsnTyrProSer----- 905
QY 1732 TGTGATATGAAGAACCGGTGTTTAAACCCCTTATGTGAACCCAGTTTATGACCCGATACC 1791
Db 906 -----LeuLysSerGlnLeuGlnLeuIleLeuAsn----- 915
QY 1792 TTAGCT---AAATGGGATATTGGAAGCAACAGATAGATAGATGTTGTACTGATCAT 1848
Db 916 LeuGlyAspLysLeuLysAlaIAsnAsnAspPheIleAspAspThr---ValValAsnAla 934
QY 1849 AAGAAATATGCATATGGAAGTGAAATGGAAGATGAAATTAATGCTGCTGCTATACCGGAA 1908
Db 935 LeuGlnTyrLysThrSerPheLysSerThrLeuLysValAsnSerLeuLysIleProIle 954
QY 1909 AACGCGTTTGATACACAGCGTCGATTTTGAACCTTTGTACGTGAACAATCTTTGACGGT 1968
Db 955 AsnLeuPhePhePheThrLeu---TrpLeuLysPheAsnLeuGlnIleProIleAspIly 973
QY 1969 GCCATTATTGAAACACATTAAGATCTATTAATGACCAAGTACAAATATCGATATACCG 2028
Db 974 SerLeuThrLeuThrSerValAsnVal-----ValPhePro 985
QY 2029 TCTAAACTGAAATGTGATGTGTAATGATATGATGAAATATTACTGATGAACTTAAT 2088
Db 986 TyrSerLeuTyrAspThrSerSerAsnGluPheThrArg---IleValAspArgLeuAsn 1004
QY 2089 ATG--AAACGTGATTTATTAATTAAAGACAGCGTAGAAGAAATTTCC 2130
Db 1005 PheThrAspThrAsnAsnPheThrLeuLysAspAlaPheProAsnPhe 1019

RESULT 13
US-09-627-376-7
; Sequence 7, Application US/0962376
; Patent No. 6342385
; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia Caulfield, Page Chen, Ping
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: UAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-627-376-7

Alignment Scores:
Pred. No.: 5.95e-07 Length: 990
Score: 163.50 Matches: 176
Percent Similarity: 33.73% Conservative: 136
Best Local Similarity: 19.03% Mismatches: 330
Query Match: 4.02% Indels: 283
DB: 4 Gaps: 44

US-09-727-892a-2 (1-2286) x US-09-627-376-7 (1-990)
QY 79 TACAAATGAAGTAAACGAGCAAAAAAACCAACCAATATTAACACGTTACTTATCTGTA 138
Db 80 TylaAlaArgTyrAsnThrArgSerThrProPhe-----GlyLeuPheSerSerIle 96
QY 139 GCAAATGCTTGCTTATATGCT-----TATGAATATGATGTT 174
Db 97 GlyLeuLysGlyLeuPheSerAsnHisProArgLysGlnLysSerCysTyrGlnLysSerVal 116

```

```
Db 463 PheHisSerIysLeuProAsnThrIleValThrLysAsnValAsnLysThrLysGluIle 482
      ::::: |:::::|
      ::::: |:::::|
QY 1069 TTTTACGACACTATTACAGACCAACAGCTTAATCCCTACTTTTGTAGATGACAAATTAT 1128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 PheThrGluIaIaIy-----ProAsnThrIleIleThrGluIleuAsnGluValAlProTyr 500
      TTT-----
      |||
QY 1129 TTT-----TCATTA----- 1137
      |||
Db 501 PheGlyArgGlyIaAsnIleMetIleSerAsnSerIleuLysSerHisGluIleu 520
      |||
QY 1138 -----TATACATTGATTAAGATGTATTTAACAGATTATTATTAATAATTAATCA 1191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 ArgAsnThrThrThrLysLysGluMetSerIleAsnSpleIeThrValArgAlaIaIaThrSer 540
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1192 CGGTATTA-----CGTCAATGATTGTCAAAATCTTAATTAAT 1230
      |||
      |||
Db 541 GluGluLeuThrPheTyrSerLysLysTyrGluLysArgValIlePheValMetAsn 560
      |||
QY 1231 GATATGATTACGTTATATCAATACAAATACATTAGAATGATTCAAGACATTACGGGT 1290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 MetPheAsnTyrIleAsn---GlySerLysLeuLeuArgPheLeuLeuGluValSerAsn 579
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1291 ATTATATGATGATACATACGTTGTTAAATTCGTTGTTATATATGATGATGATCTTCAT 1350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 SerAspPheGluAsnIleThrProIleThrLeuGlySerLeuAsp---SerTyrAsnHis 598
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1351 GCACGTGATATTTTTCAAAATCTTTTAT 1383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 ValProAlaIleIleTyrLysAspIleIleIleLysProGluThrTrpAsnIleArgLys 618
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1384 -----AAACACCAAGGTAAAGTAAACAAATC-----AATAGACATCCACT 1428
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 SerGluAlaLysThrLeuAspSerLeuLysAsnThrPheuAsnAsnValProPro 638
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1429 TAC-----GACTATCATTACGATGAT 1452
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 639 PheValArgMetLysTyrThrAspGluIleIleTyrLeuAspLeuSerArgThrIleAsp 658
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1453 -----ATCACGACACCCCATAC----- 1470
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 LeuThrMetLeuPheGlnSerIleLysLysHisSerPheIleGlnLeuLeuAspValHis 678
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1471 -----TCAATGACGAGCTTATGTTATCTTAAGTCGTTTAAATGATTAATAGGC 1521
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 679 SerValCysThrAsnAspThrGluIleLeuGluValVal----- 692
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1522 ATACCTGCATTAACGTTCA-----CATTT-----AACTTA 1551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 ---ProPheThrArgSerAspValAsnAlaHisGlnIleTyrHisTyrAlaGluAsnIle 711
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1552 TTCGCTTAGATGAT-----AACATGACCTATACATATACATTAACGCTTACAAAAA 1605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 TyrThrLeuGluAspSerGlySerLysGluLysTyrPheTyrAlaLysIleTyrValAsn 731
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1606 ACTGACGTAATATATTAAT----- 1626
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 732 LysGlnArgGluThrSerPheLeuGlnLysGluTyrProLeuLeuLeuLysTyrLeuLys 751
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1626 ----- 1626
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 752 LeuProGluAsnLeuGlnTrpPheTyrIleArgTyrLysAspAspGlyLysAspSerIle 771
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1627 -----TCTACATTGTCATCAGCTTCATGTTATTAATCTTATGTTCCCTTC--- 1674
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 ArgLeuArgIleArgTyrValGluAspLysGluLeuValGlnLeuTyrSerArgPheIle 791
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1675 -----CAATAC 1680
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 GluTrpAlaThrLysAlaLysAsnIleGlnIleSerGlyTyrGluIleSerGluTyr 811
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1681 TTAACCGAA-----AGTGAATTCGACGACATTTTAT 1713
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 812 IleProGluSerAlaArgTyrGlyLysLysTyrSerSerIleIleHisSerPhePhe 831
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1714 TATTCGGATAC-----ATAGTTGTGTATATGAAG-----TCCGTTGTTAAACC 1758
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 TyrTyrAspSerIleLeuAspLeuLeuLeuGlnLysLysAlaGluIaIaIleGluVal 851
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1759 TATATGACCCCGATTATTCGACCCGATACGCTTATAGTAAATGGGATATGAACGAA 1818
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 ArgThrSerLeuSerIleIleArgMetPheLeuMetIleLysLeuSerGlnAspGln 871
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1819 CAG-----ATACATTAAGATGTTGTACTGATCATCAAG---AAATATGCATATAGAGTG 1869
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 872 GlnLysLeuIleLysAsnLeuPheAspGlyLysHisLysLeuLysTyrGluLysGluTyr 891
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1870 AATGAAAGATTAAATTTGCTTCTGCTGTATA---CCGAAAAACGCTTTGATACACG 1926
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 HisAsnSerIleSerLeuLeuLeuAspAsnLeuLysThrLysAsnGlnIaIaIaIa 911
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1927 GTTCGATTTGAAACCTTTGTACGTGACAAATTTTGTACGCGTCCATTATGAAAACAAT 1986
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 AspIlePheCysValMetAsnMetLysLys-----IleThrGluLysIle 926
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1987 AAAGATATCTATATGAGCAAGGTACATATCG-----ATATATCCGTCTAAA 2034
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 927 SerSerValLeuLysGlnLysAspLeuThrThrAspTyrPolaArgIleLeuGlySerIle 946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2035 ACTGAATGTGATGTGTGTAATGATATGATGAAATATTACTGATGACTTAATATGAAA 2094
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 IleHisMetArgCysAsnArgValTyrGly-----IleAsnSerGluLeuGluArgLys 964
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2095 CGTGATTTATATTA 2109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 ThrMetPheIleVal 969
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-134-001C-3242
; Sequence 3242. Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3242
; LENGTH: 1151
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3242

Alignment Scores:
Pred. No.: 7,72e-07 Length: 1151
Score: 162.50 Matches: 165
Percent Similarity: 35.13% Conservative: 135
Best Local Similarity: 19.32% Mismatches: 339
Query Match: 3.99% Indels: 215
DB: 4 Gaps: 42

US-09-727-892a-2 (1-2286) x US-09-134-001C-3242 (1-1151)
QY 214 TATACGTATGTGAAGAACGTGATACATC-----ACAAATGCAAAAACGATATT 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 TyrThrLysLeuLysThrTyrAsnAspValLysGlnMetAsnArgGluAsnValAspVal 122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 265 ATCATGATTCGACATATAC-----GTAAATAAATACGATATCATTTTCTTACTTAAGAC 318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 GluSerIleGluIleAsnPheProSerAsnLysPheAsnLeuGlyPheThrPheGluAsp 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 319 ACCATGCGTTATTTGATTAATTTACACCGCAAAATATATTTAAATCTGCAGAGAA 378
    |||
Db 143 ThrGlyValAspSerAsnValAla-----ThrHisGlnSerSerThrGln 159
QY 379 AATGACACACATTTAAATTAAGAGGCTACTATTTAGCCCAAAATCAAATGTAAAT 438
    |||
Db 160 PheMetLysThrSerAsnLeuLeuPheLysThrVal-----AspTyrAsn----- 174
QY 439 TTGAAAACCGCTTAAATCTTCATCATTTAGATTAAACAATGTTTTAAATGCT--- 495
    |||
Db 175 -----HisValGlnSerAlaLeuAsnPheLysPheMetLysArgIleAsnGluVal 191
QY 496 -----TTAAATTTAAATATTGTAACCTT----- 522
    |||
Db 192 GlyIleProIleLeuPheValIleAsnGlnIleAspLysHisAsnGluGluIleThr 211
QY 523 -----ATGAAACCAATACATTCATTCACATTTAGTAGAATTAATCTT 567
    |||
Db 212 PheGluThrPheLysSerArgValGluLysSerIleLysAspThrPhePheLysLeuGln 231
QY 568 GATGGTGTTATTTAACA-----GAATCACAACCTTAAACAGATTTTAAT 612
    |||
Db 232 AspThrTyrTyrValSerLysPheAspHisProGlnAsnGluIleAspLysLeuSerAsn 251
QY 613 TATACGATTTTGTATAAGATTAATGATGATAGTGAACCGCTATGCTGCTG 672
    |||
Db 252 PheLeuValPhe-----MetAspGlnHisArgGlnSerThrGluAspTyrValAsn 268
QY 673 AAATGTTTCCAAACTCACACCTGGAACAACCTTACATCATCATTAAGACGCG---ATT 729
    |||
Db 269 ArgThrIleGlnPheIleThrAspAlaGlnTyrIleTyrIleGlnAsnGluMetGlnSer 288
QY 730 AATATTAGTATGTCCTATTCAT-----TATAGTATATATTTCCAAATTTTGAC 780
    |||
Db 289 IleLeuAspThrLeuGlnIleAsnGluGluGlnPheGluGluAlaTyrIleGlnPheGln 308
QY 781 TATACAAATTAACATTTTCATTAATGATTAATGATCTTACTTACATTAATAAGAC 840
    |||
Db 309 GlnAsnGln-----GluValSerAlaGluAlaGlnLeuLeuAsnAspSerAsnGln 325
QY 841 CGTTTCAGTACTCAACCATATCAAGATTTAAATTTCTTAATACATTAATCATTTTC 900
    |||
Db 326 LeuPheAsnTyrLeuLysGlnLysArgLysAspIleLeuAspAsnAlaTyrIleMetThr 345
QY 901 CATGATATG-----AATTTTATGACTATATTAAATCA-----TTCTATGCT 942
    |||
Db 346 TyrAspMetArgLysLeuSerLeuArgAsnTyrLeuGlnSerMetAlaThrAspPheLysVal 365
QY 943 GGTGGT-----TTAAATATG 957
    |||
Db 366 AsnGlyPhePheAsnLysArgLysLysLysGluGluGlnIleLysArgLeuAsnGlu 385
QY 958 TATACACCAAAATAC-----ATAACCAAACTAATGATGAGCGCTGTTTCTATT 1008
    |||
Db 386 AlaThrThrGlnLeuGlnIleLysValAsnGlnGlnValArgGlnIlePro--LeuArgGlu 404
QY 1009 GACATCAATTCAGATTATCTTATGATGATCAAT-----GAAAAATTCACACA 1059
    |||
Db 405 AspMetSerPheLeuThrArgPheIleAsnLysHisAlaValAsnGluLysIleLeuAsn 424
QY 1060 TGGTTACTTTTACCAACATTCAGACCAACAGCTTAATCCCTACTTTTAAATGAT 1119
    |||
Db 425 -----GlnGluTyrAspValAlaProSerLeuIleSerGlu 436
QY 1120 GACAAATTTATTTATATATAGATTGATAAGATGATTAAAC----- 1164
    |||
Db 437 -----LeuTyrGlnThrGlnThrSerIleSerAspThrTyrValLeuThr 451
QY 1165 ---GATGATTTTAAATTAATTAATCAACGCTGATTATGCTCAAAATGTTAAATAC 1221
    |||
Db 452 PheSerAspGluValIleLysAlaLeuAsnLysLysIleGluAsnGlnSerThrProLeu 471
QY 1222 TATATAATGATTAATGATTAACGTTATATATCAATTAACAATTAAGATTAATGAC 1281
    |||
Db 472 PheGluGluAlaValAsnHisValGlnValAsnGluLeuSerSerAspGluAsnGluAsp 491
QY 1282 ATTACGGGATTCGATTGATGCATATATACGTTAAATCGTTGCTT----- 1326
    |||
Db 492 ArgTyrGluTyrAsp---ArgTyrIleGluLeuAsnThrLeuLysAspSerLeuThrSer 510
QY 1327 ---ATATGATATGTCATATCTTCATGCAACCGTGAATATATTTTCAAAACATTTTAA 1383
    |||
Db 511 HisAspTyrLysHisTyrTyrIleHisLeuAspAspSerLeu-----AspLysLeuIle 528
QY 1384 -----AAAAACAAGGTAAATTAAAAACAATCAATATGACATACACCTTAACGAC 1434
    |||
Db 529 GlyArgThrGluThrHisPheGluLeuLysGlnGlu--AsnSerThrAlaTyrHisArg 547
QY 1435 TATCAATATACGATGATATTCACAGACACCCATCTCAATAGACAGGTTATGTTATCT 1494
    |||
Db 548 LysHisGluThrGlnHisArgAsnGluPheValIleThrSerAsnGlnAspIle----- 564
QY 1495 AAGTCGTTTAAATGATTAATGATATATGACATACCTGCATTAACGTTCAATTAATCTTTC 1554
    |||
Db 565 LysArgAlaLeuAspIleValLysAspValProLeu-----PheAspArgThr 580
QY 1555 CGTTTAGATGATTAACAATGAACATATACAAATATCAATTAACGCTTCAAAACAACCTGAACGT 1614
    |||
Db 581 LysGlnAspIleThrAspThrIleLeuArgLeuAspAsnGlnIleThrLysValGly--- 599
QY 1615 AATATATTTCTCTACATTTGTCAATCAGTCTCAATCAGTCT---TCATGTTAATCTTGGTCTCT 1671
    |||
Db 600 -----ValPheGlyThrPheSerAlaGlyLysSerLeuIleAsnAlaLeuLeuGly 617
QY 1672 TTCCAATACCTTAACG----- 1686
    |||
Db 618 AspAsnTyrLeuValSerSerProAsnProThrAlaAlaThrThrGluLeuSerTyr 637
QY 1687 ---GAAGTGAATTTGAC----- 1701
    |||
Db 638 GlyLysGluSerGlnIleThrLeuLysSerLysGluGlnLeuGluGluValAsnHis 657
QY 1702 -----GACAAATTTATTTATGCGATCT 1725
    |||
Db 658 ValLeuGluPheTyrGluIleSerPheAsnThrLeuAspAsnPheIleLysSerAspLeu 677
QY 1726 GATACTTTGTTATGAA----- 1743
    |||
Db 678 GlnLysLeuLysSerLysLeuGlnLysAsnGlnLeuAlaPheIleSerAlaIleGluLys 697
QY 1744 ---TCCGTTGTTAAACCTTATTGAACCCGAGTTTATTCGACCGATGCTTA--- 1794
    |||
Db 698 HisTyrGluMetTyrThrSerMetLeuGlnHisSerLeuIleHisThrValSerLeuGlu 717
QY 1795 ---GGTAATGGGATATGAAACGACAGATAGATTAAGATGTTGTACTGAATCAT 1848
    |||
Db 718 GlnIleLysLysThrPheSerAlaGlnAspGluTyrAlaThrPheValLysThrValHisLeu 737
QY 1849 AAGCAATATGATGATGAAGTGAAGGAAGATTAATTTGGTCTGCTGGATACCCGAAA 1908
    |||
Db 738 LysLeuProLeuAspThrPheLysGlyLysIleIleIleAspSerLeuGlnLysHisSer 757
QY 1909 AAC-----GCTTTGATACAAACGCTGAT----- 1932
    |||
Db 758 AsnAsnGlnArgHisThrAsnGluThrGluGlnIleLeuThrSerSerAspLeuIleLeu 777
QY 1933 TTGAAACCTTTGATGACGAACAATCTCTTGACCGT-----GCCATTTATGAACAAT 1986
    |||
Db 778 TyrValThrTyrThrPheAsnHisSerPheThrAspAsnAspLysAlaPheIleGluHisMet 797
QY 1987 AAAAGTATC-----TATAATGAGCAA-----GCTAACATTTGCTG 2019
    |||
Db 798 LysAspMetAsnGlnLeuAsnGlnLysAlaPheLeuPheLysMetIleIleAsnAlaValAsp 817
QY 2020 ATATATCCGCTTAAACTGAATTTGATGTGATGATGATGATGATGATGATGATGATGAT 2079
```

Db 818 leuAlaIuaSpLysGlnAspIle-----GlnAlaValGlnAspTyrValAlaAsp 834
QY 2080 -----GAACCTTAATATGAACGTGAATTTATTAAGACGCTAGAGAAATTC 2130
Db 835 AlaLeuGlyGlnValAsnLeuHisSerSerPileTyrSerValSerSerArgInSerLeu 854
QY 2131 GACCATAGTCATTT-----GATGATATTTCTTTATTTGAAGTAC 2172
Db 855 AsnGlyAsnAsnIleGlyIleAsnGlnLeuArgGlnSerIleGlnTyrPheAlaVal 874
QY 2173 ATCGGTCATTTTCACTTAACGACTTATTCACGTTGACGCTTAGTACATACAAATCT 2232
Db 875 GluSerArgThrIleLeuGlnGlnGlnIleMetThrTyrGlnLeuGlnIleMetAsnAsnSer 894
QY 2233 GATTGATATATTAACGTGAACATGATGAATTAATAAA 2274
Db 895 PheLysLysMetIleLysAspPheHisAspAsnAlaLys 908
RESULT 15
US-08-480-604A-10
Sequence 10, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BROCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
Application Number: US/08/480,604A
Filing Date: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
Application Number: US 08/4422,711
Filing Date: 14-APR-1995
PRIOR APPLICATION DATA:
Application Number: US 08/405,496
Filing Date: 16-MAR-1995
PRIOR APPLICATION DATA:
Application Number: US 08/329,154
Filing Date: 25-OCT-1994
PRIOR APPLICATION DATA:
Application Number: US 08/161,907
Filing Date: 02-DEC-1993
PRIOR APPLICATION DATA:
Application Number: US 07/985,321
Filing Date: 04-DEC-1992
PRIOR APPLICATION DATA:
Application Number: US 07/429,791
Filing Date: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-10
Alignment Scores:
Prod. No.: 1.2e-06 length: 2366
Score: 161.50 Matches: 169
Percent Similarity: 33.77% Conservative: 115
Best Local Similarity: 20.10% Mismatches: 306
Query Match: 3.97% Indels: 251
DB: Gaps: 46
US-09-727-892A-2 (1-2286) x US-08-480-604A-10 (1-2366)
QY 196 TCTTTTATGCGCATTTTATACGATGTGAAAGAGTGAT----- 237
Db 1315 SerPheTyrGlySerGlyGlyThrTyrAlaLeuSerLeuSerGlnTyrAsnMetGlyIle 1334
QY 238 ACAATCACAATAATCAAAACAGATATTATCATGATGACATACGTATAATAATACAT 297
Db 1335 AsnIleGlnLeuSerGlnSerAspValIlePheIle----- 1346
QY 298 AATCATTTTACTTAAGACACCATCGCTTATTGTATATATATACACGCGAAATATA 357
Db 1347 -----AspValAspAsnValAlaArgAspValIleThrIleGlnSerAsp 1360
QY 358 TATTTAAATCTCAGAGAAATGCA-----CACACATTAATAATGAAGCGCT 408
Db 1361 LysIleLysLysGlnAspLeuIleGlnGlyIleLeuSerThrLeuSerIleGlnGlnAsn 1380
QY 409 ACATTTTACCAAAATCAAAATGTAATTTTGAAGAAACGTGTAATCTTCATCAAT 468
Db 1381 LysIleLeuAsnSerHisGlnIleAsnPheSerGlyGlnValAsnGlySer----- 1398
QY 469 TTGATTTTAACAATGTTTAAATGCTTT-----AAATTATATATATGATAC 519
Db 1399 -----AsnGlnPheValSerLeuThrPheSerIleLeuGlnGly 1411
QY 520 TTATGAAACAAATCATCATTCATGCAACATTCAGTAC-----AAATTACTGATGT 573
Db 1412 ---IleAsnAlaIleIleGlnValAspLeuLeuSerLysSerTyrLysLeuIleSer 1430
QY 574 GGT-----TATTTAACAAGATCAAACTTAAACACATTTAATTATACG 618
Db 1431 GlyGlnLeuLysIleLeuMetLeuAsnSerAsnHisIleGlnGlnLysIleAspTyrIle 1450
QY 619 ATTTTGATTAAGATATATGATATGAAT-----GATAGTGAAGCTAT 666
Db 1451 GlyPheAsnSerGlnLeuGlnLysAsnIleProTyrSerPheValAspSerGlnGlyLys 1470
QY 661 GACTATGCTGTGAATGTTTGCAGAACTCACACCTGACACACTTACATCATTCATAAT 720
Db 1471 GluAsnGly-----PheIleAsnGlySerThrLysLysGlnGlyLeuPheValSerGln 1487
QY 721 -----GACGTGATATATATAGTATGTCGCAATTCATATGATATATTTGCAAT 774
Db 1488 LeuProAspValAlaLeuIleSer-----LysValIleThrMetAspSerLysProSer 1505
QY 775 TTGACTAT-----AACAAATTA-----ACATTTTCAATGAAT 807
Db 1506 PheGlyTyrTyrSerAsnAsnLeuLysAspValIleThrLysAspAsnValAsn 1525
QY 808 ATATAGCAATCTTACTGAAATATGAATGACACGTTTCACTTACCAACATATCA 867
Db 1526 IleLeuThrGlyTyrTyrLeuLysAspAspIleLysIleSerLeuSerLeuThrLeuGln 1545
QY 868 GATATTAA---ATATCTTATACACATTTATCATGATATGATATTTTATGACTAT 924
Db 1546 AspGlnLysThrIleLysLeuAsnSerValHisLeuAspGlnLeuSerGlyValAlaGlnIle 1565

QY 925 ATTAATCATCTTATCGTGGT-----TTA 951
 Db 1566 LeuyspheketaanaaglyglnthrasnthSeaspsSerLeuKetSerPheLeu 1585
 QY 952 AATATGATTAACACCA-----TACATAACAACTAATGATGACGCTGTTTTCG 1005
 Db 1586 GluSerMetasnileysSerilePheValasnPheLeuGlnSerasnileysPheile 1605
 QY 1006 ATTGACATCAAT----- 1017
 Db 1606 LeuaspAlasnPhelileSerGlyThrThrSerileGlylnPheGlnPheileCys 1625
 QY 1018 -----TCGAGTTATCCTTATGTGATGATGATCAATAAATAATTCACATGTTA 1065
 Db 1626 AspGlnasnspasnileglnProTyrPheileLysPheasnThrileGlnThrAsnTyr 1645
 QY 1066 TACTTTACGACACTATTGACACCAACCTTAATC---CCTACTTT---TTAGATGAT 1119
 Db 1646 ThrLeuTyrValglnasnarglnasnMetileValGlnProasnTyrAspLeuasp 1665
 QY 1120 GACATATTATTTCA-----TTATATAAGATTGAT 1149
 Db 1666 SerGlyAspIleSerSerThrValileasnPheSerGlnLysTyrLeuTyrGlyileasp 1685
 QY 1150 AAA-----GATGATTTAACGATGATTTATTAATAA 1182
 Db 1686 SerCysValasnLysValValileSerProasnileTyrThrAspLulleasnileThr 1705
 QY 1183 -----ATTAATCACCTGTATTACGCAATGATGTGTAATAACATATAATGAT 1233
 Db 1706 ProValTyrGlnThrAsnAsnThrTyrProGlnValileValLeuaspAlasnTyrile 1725
 QY 1234 AATGATACCTTAATATCAATCAATAAAT-----ACATTAAGATGATTAACAGCATACG 1287
 Db 1726 AsnGlnLysileasnValasnileasnAspLeuSerileargTyrValIrrpsAsnasp 1745
 QY 1288 GGTATTCATTCATGATATACGTGTTAATTCGTTTGTATATAT-----GATGT 1338
 Db 1746 Gly-----AsnaspPheileLeuMetSerThrSerGln 1757
 QY 1339 GATATCTTCATGCAGTATATTATTTTCAAACTATTTTAAACACAGGTAG 1398
 Db 1758 AsnLysValSerGlnValLysileargPheValasnValPhe-----LysaspLysThr 1775
 QY 1399 TTAATAAACAAATCAATATGACATCACCTTACGACTACATTAAGTATATCAAC 1458
 Db 1776 LeuAlasnLysLeuSer-----PheasnPheSerasp 1786
 QY 1459 GACACCCCACTCAATGAGAGGTATGTTATCTAAAGTCGTTTA----- 1506
 Db 1787 -----LysGlnaspValProValSerGlnLulleLeuSerPheThrPro 1801
 QY 1507 -----AATGATTATATGACATACCTGATTCAGTTACATTTTAACATTATTC 1554
 Db 1802 SerTyrTyrGlnaspLysPheileGly-----TyraspLeuGly 1814
 QY 1555 CGTTTGATGATACATACATATACATATACCGTTACAAAACACTGACGT 1614
 Db 1815 LeuValSerLeuTyrasnGlnLysPhe-----TyrileasnasnPheGly 1829
 QY 1615 AATATATTTATCTCTACATTTGACATCACGTTTCATTTGATTAATGTTGTTTC 1674
 Db 1830 MetMetValSerGlyLeuileTyrileasnAspSerLeuTyrTyrPheLysProProVal 1849
 QY 1675 CAATACTTAACGGAAGTAATGACGACAAATTTATTTATTCGATACGATAGTTTG 1734
 Db 1850 AsnAsnLeuile-----ThrGlyPheValThrValGlnaspLysPheTyr 1864
 QY 1735 TATATGAATCCGTTTAAACCTTATGAAACCCGTTTATGACCCGATAGCCTTA 1794
 Db 1865 Tyr-----PheasnProIleasnGly 1871

QY 1795 GGTAAATGGGATATTGAAACACAGATAGATAGATGTTTGTACTGAATCATACAGAA 1854
 Db 1872 GlyAlaAlaSerileGlyGlnThrIleileasp-----AspLysasn 1885
 QY 1855 TATGCATATGAGTGAATGACAAAGATTAATAATGCTTCGTCGATAT----- 1902
 Db 1886 TyrTyrPheasnGlnSerGly-----ValLeuGlnThrGlyValPheSerThrGln 1902
 QY 1903 -----CGAAAAACGCTTGTATACAAAGCGTCGATTTGAAACG 1941
 Db 1903 AspGlyPheLysTyrPheAlaProAlasnThrLeuasnAspGlnasnLeuGlnGlyAla 1922
 QY 1942 TTTGACGGAACATTCCTTTGACGTCGATTTGAAACAAATAAGTATCAT--- 1998
 Db 1923 Ile-----AspPheThrGlyLysLeuileIleileaspGlnasn-----IleTyrTyr 1937
 QY 1999 -----AATGACCAAGTACAAATATCG----- 2019
 Db 1938 PheasnspasnTyrArgGlyValValGlnThrPlyGlnLeuaspGlyGlnMetHisTyr 1957
 QY 2020 ATATATCCGTCATAAAGCAATGTATGCTG-----AATGTATATGATGAA 2067
 Db 1958 PheSerProGlnThrGlyLysAlaPheLysGlyLeuasnGlnIleGlyAspTyrLysTyr 1977
 QY 2068 TATTTTACGATGAACCTTAATATGCAACGTCATTTATA---TTAAAGACGTCAGAGA 2124
 Db 1978 TyrPheasnSerAspGlyValMetGlnLysGlyPheValSerileasnAspasnLysHis 1997
 QY 2125 AATTCGACACATGACATTTGATATCTTTATATGAAAGTAC----- 2172
 Db 1998 TyrPheasnspSerGlyValMetLysValGlnTyrThrGlnLulleaspGlyLysHisPhe 2017
 QY 2173 -----ATCGGTTCATTTTACTTAACGACTTATTTCA 2205
 Db 2018 TyrPheAlaGlnasnGlyGlnMetGlnIleGlyValPheasnThrGlnaspGlyPhe--- 2036
 QY 2206 GTTGAACGTTAGTACATACAAATATGATTTGCATATATTAATAACGTCGAACATGATGAA 2265
 Db 2037 ---LysTyrPheAlaHisHisasnGlnaspLeuGly-----AsnGlnGlnGlyGln 2053
 QY 2266 ATA 2268
 Db 2054 Ile 2054

Search completed: January 8, 2003, 17:03:52
 Job time : 79.5 secs

Thu Jan 9 14:50:06 2003

us-09-727-892a-2.rapb

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: January 8, 2003, 17:06:20 : Search time 15.5 Seconds
(without alignments)

5722.686 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 4070
Sequence: 1 atggagatctagatgcat.....taaaaaagcaactgttaa 2286

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 118974 segs, 19401057 residues

Total number of hits satisfying chosen parameters: 237948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+.n2p.model -DEV=xlh
-O=/cg2_1/USPTO.spool/US09727892/rnat_06012003_151203_9434/app_query.fasta_1.2439
-DB=published.Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62
-TRANS=human40.cti -LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -REARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09727892 -CGN=1.1.7 -rnat_06012003_151203_9434
-NCPU=6 -ICPU=3 -NO.XLPXY -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications AA:

1: /cg2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
2: /cg2_6/ptodata/2/pubppa/PTCT_NEW_PUB.pep.*
3: /cg2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cg2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cg2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cg2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
7: /cg2_6/ptodata/2/pubppa/PTCTUS_PUBCOMB.pep.*
8: /cg2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cg2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
10: /cg2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
11: /cg2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
12: /cg2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
13: /cg2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
14: /cg2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	163.5	4.0	990	12 US-10-047-676A-7	Sequence 7, Appl1
2	159	3.9	2167	10 US-09-801-368-56	Sequence 56, Appl1
3	156.5	3.8	476	10 US-09-774-414-3	Sequence 3, Appl1
4	155	3.8	824	9 US-10-011-588-33	Sequence 33, Appl1

5	152.5	3.7	1427	10 US-09-801-368-354	Sequence 354, App
6	149.5	3.7	773	12 US-10-067-385-8	Sequence 8, Appl1
7	148	3.6	897	10 US-09-815-242-11609	Sequence 11609, A
8	145	3.6	872	9 US-09-843-676-8	Sequence 8, Appl1
9	145	3.6	872	9 US-09-843-676-54	Sequence 54, Appl1
10	145	3.6	872	9 US-09-766-253-8	Sequence 8, Appl1
11	144.5	3.6	872	9 US-09-766-253-54	Sequence 54, Appl1
12	144.5	3.6	745	10 US-09-815-242-12209	Sequence 12209, A
13	143.5	3.5	1295	10 US-09-726-949A-1	Sequence 1, Appl1
14	142	3.5	972	9 US-09-879-955-10	Sequence 10, Appl1
15	136	3.3	996	10 US-09-815-242-5251	Sequence 5251, Ap
16	136	3.3	1009	10 US-09-815-242-12141	Sequence 12141, A
17	135	3.3	592	9 US-09-286-488-24	Sequence 24, Appl1
18	135	3.3	592	10 US-09-737-118-24	Sequence 24, Appl1
19	135	3.3	929	10 US-09-815-242-12273	Sequence 12273, A
20	131.5	3.2	1781	9 US-09-995-749A-2	Sequence 2, Appl1
21	131	3.2	866	10 US-09-815-242-11373	Sequence 11373, A
22	131	3.2	929	10 US-09-815-242-12903	Sequence 12903, A
23	130.5	3.2	917	9 US-10-001-835-189	Sequence 189, App
24	130.5	3.2	993	10 US-09-815-242-5809	Sequence 5809, Ap
25	130.5	3.2	1002	10 US-09-815-242-12899	Sequence 12899, A
26	130.5	3.2	1002	10 US-09-815-242-13158	Sequence 13158, A
27	128.5	3.2	811	9 US-10-011-588-29	Sequence 29, Appl1
28	128.5	3.2	852	9 US-10-011-588-25	Sequence 25, Appl1
29	128.5	3.2	858	9 US-10-011-588-41	Sequence 41, Appl1
30	128	3.1	809	9 US-10-105-695-2	Sequence 2, Appl1
31	128	3.1	809	9 US-10-105-694-2	Sequence 2, Appl1
32	128	3.1	809	10 US-09-747-521-2	Sequence 2, Appl1
33	128	3.1	809	12 US-10-106-014-2	Sequence 2, Appl1
34	128	3.1	3092	10 US-09-801-368-172	Sequence 172, App
35	127	3.1	1088	9 US-09-423-126-1	Sequence 1, Appl1
36	127	3.1	1088	10 US-09-280-197-1	Sequence 1, Appl1
37	127	3.1	1091	9 US-09-423-126-2	Sequence 2, Appl1
38	127	3.1	1091	9 US-09-280-197-2	Sequence 2, Appl1
39	123.5	3.0	1007	9 US-09-843-676-86	Sequence 86, Appl1
40	123.5	3.0	1007	9 US-09-766-253-86	Sequence 86, Appl1
41	123.5	3.0	1031	9 US-09-843-676-2	Sequence 2, Appl1
42	123.5	3.0	1031	9 US-09-766-253-2	Sequence 2, Appl1
43	123	3.0	924	10 US-09-816-028A-10	Sequence 10, Appl1
44	123	3.0	1040	10 US-09-864-761-38325	Sequence 38325, A
45	123	3.0	1041	9 US-09-978-295A-498	Sequence 498, App

ALIGNMENTS

RESULT 1
US-10-047-676A-7
: Sequence 7, Application US/10047676A
: Patent No. US20020123105A1
: GENERAL INFORMATION:
: APPLICANT: Qi, Fengxia
: APPLICANT: Caulfield, Page W.
: APPLICANT: Chen, Ping W.
: TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
: FILE REFERENCE: UAB-17403/22
: CURRENT APPLICATION NUMBER: US/10/047,676A
: PRIOR FILING DATE: 2002-03-21
: PRIOR FILING DATE: 2000-07-28
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: Patentin version 3.1
: SEQ ID NO 7
: LENGTH: 990
: TYPE: PRT
: ORGANISM: Streptococcus mutans
US-10-047-676A-7

Alignment Scores:
Pred. No.: 1.24e-06
Score: 163.50
Percent Similarity: 33.73%
Best Local Similarity: 19.03%
Query Match: 4.02%
Length: 990
Matches: 176
Conservative: 136
Mismatch: 330
Indels: 283


```
Db 752 LeuProGluAsnLeuGlnTrpPheTyrIleArgTyrIleAspAspGlyLysAspSerIle 771
QY 1627 -----TCACATTGTCACATCACGTCCTTACATGATTAATGTCCTTC--- 1674
Db 772 ArgLeuArgIleArgTyrValGluAspLysGlnLeuValGlnLeuTyrSerArgPheIle 791
QY 1675 -----CAATAC 1680
Db 792 GluTrpAlaThrLysAlaArgLysAsnIleGlnIleSerGlyTyrGluIleSerGlyTyr 811
QY 1681 TTAAACGAA-----AGTGAATTCGACACAAATTTTATT 1713
Db 812 IleProGluSerAlaArgTyrGlyLysLysTyrSerSerIleIleHisSerPhePhe 831
QY 1714 TATTCGATCTCT-----GATAGCTTTGTATATGAAA-----TCCGTTGTTAAACC 1758
Db 832 TyrTrpAspSerIleLeuAspLeuLeuLeuGlnLysAlaGluGlnThrIleGluVal 851
QY 1759 TTATTGACCCCGCATTTATTCGACCCGATAGCCCTTAGCTAATGGATATTGAAACGAA 1818
Db 852 ArgThrSerLeuSerIleIleArgMetPheLeuMetMetLysLeuSerLeuGlnAspGln 871
QY 1819 CAG-----ATAGATAGAGCTTTGCTACTGATCATTAAG---AAATATGCATATGAAGTC 1869
Db 872 GlnLysLeuIleLysAsnLeuPheAspGlyLysHisLysLeuLysTyrGluLysGlyTyr 891
QY 1870 AATGGAAGATTAAATTTGCTTCGCTGTATATA---CCGAAAAACGCTTTGATACACAGC 1926
Db 892 HisAsnSerIleSerLeuLeuLeuAspAsnLeuLysThrLysAsnGlnThrAspGluAla 911
QY 1927 GTTCATTTTGAACCTTTGATAGCGAACAAATCTTTAGCGGTGCCATTATTGAAACAAAT 1986
Db 912 AspIlePheCysValMetAsnMetLysLys-----IleThrGlnLysIle 926
QY 1987 AAAAGTATCTATTAATGAGCAAGTACAAATATCG-----ATATATCCCTCTAAA 2034
Db 927 SerSerValLeuLysGlnLysAspLeuThrThrAspTrpGlnArgIleLeuGlySerLeu 946
QY 2035 ACTGAATTTGATGTGTAATGTATATGATGATGAAATTTTACGTAGCACTTAATATGAAA 2094
Db 947 IleHisMetArgCysAsnArgValTyrGly-----IleAsnSerGlnLeuGluArgLys 964
QY 2095 CGTGAATTTATATTA 2109
Db 965 ThrMetPheIleVal 969

RESULT 2
US-09-801-368-56
; Sequence 56, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OR INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 2167
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-56

Alignment Scores:
Pred. No.: 3.5e-06 Length: 2167
Score: 159.00 Matches: 167
Percent Similarity: 33.61% Conservative: 118
Best Local Similarity: 19.69% Mismatches: 293
Query Match: 3.91% Indels: 270
DB: 10 Gaps: 40

US-09-727-892a-2 (1-2286) x US-09-801-368-56 (1-2167)
QY 205 GACGACATTTTATACGATGTGAAAAGACGTATACATCAAAATCAAAACGATATT 264
Db 1061 AspPheAsnAsnIleGlyGlnValSerAspSerLeuLysGlnSerIleIleArgVal 1080
QY 265 ATGATGATTCACATACATCTATTAATACGATTAATCATTTTCTACTTAAGCACCATG 324
Db 1081 ValLeuLysSerSerLeuGlnLysIlePheAspLeuValLeuThrSerAsnIle 1100
QY 325 -----CGTATTTTGTATATATTACACGCGAAATATATATTAAATCTCGCAGAGAA 378
Db 1101 PheSerLysLeuValAspThrLysAspLeuGluAsnTyrTyrHisGlnArgGlnArg 1120
QY 379 AATGACACACATTAATAAATGAAGAGCTACTATTTTACCCAAAATCAAAATGTAAT 438
Db 1121 GlyHisSerThrArgGlyLeuSerAsp-----AspAsnIleGly 1133
QY 439 TTAGAAAACGCTTAATCTTCATCAATTTTACATTTTAAATGCTTTT 498
Db 1134 LeuLeuAspTyrLysAlaPheValLysLeuThrMetAspAsnAspIlePheThrGluThrPhe 1153
QY 499 AAATTAATATATTGATGACTTTATGAAACCAATACATCAATTCACATTAAGGTAAG 558
Db 1154 ---PheAsnThrTyrLysSerPheThrThrThrValLeuGluAsnMetAlaLys 1172
QY 559 -----AAATTACTGTGATGCTTAT 579
Db 1173 ArgTyrValGlyAlaLysSerCysSerValSerIleSerLysIleLeuAsp----- 1189
QY 580 TTAACAGATACACAACTTAACAGATTTTAATTAATGATTTTGTATAAGATTAATGAT 639
Db 1190 -----ArgSerAspSerLysMetLysIleAsnGluAspThrAsn 1203
QY 640 ATGATGATGATGAGCGCTATGACTATGCTGTAATATGTTT---GCAAAACACACACT 696
Db 1204 LeuValSerSerSerLeuTyrAspGlnAsnPheProValIleTrpMetLysValThrAsp 1223
QY 697 GAACAA-----CTTACATACATTCATTAATGAGTGAAT-----ATA 732
Db 1224 AspGlnAsnIleAsnLeuIleTyrMetAlaLysIleGlnIleGlyAlaIleGluAlaIle 1243
QY 733 TTACGTATGTCGCATATTCATTTAGATATATTTCCAAATTTTGACTATAACAA--- 789
Db 1244 LeuHisLeuValLysAsnHisTyrSerAspPheThrAspAspLeuCysAsnAsnSerThr 1263
QY 790 TTACATTTTCATGATATTAATGAACTTACTGATGATATGAAATG---ACACGTTT 846
Db 1264 LeuLeuAspIleIleLysIleMetGlnGlnValSerThrGluTrpProThrArgGlyIle 1283
QY 847 -----CAGTTA 852
Db 1284 AlaAsnSerLysLeuGlnLysSerLeuProGluAsnPheValIleGluTrpGluAsnLeu 1303
QY 853 CTCACCAATATCAAGATATTAAATATCTTATACACATATATCATTTCCATGATGAAT 912
Db 1304 LeuThrThrLeuThrAspLeu-----PheHisGlyIleLys 1315
```

QY 913 TTTATGACATATTAATCATCTCATCTG----- 942
 Db 1316 ---SerIatyrGlnLysGlnLeuItyrArgProIleGlyValAsnArgThrGlnLysArg 1334
 QY 943 -----GGTGGTTAAATATGTTAACAC-----AAATACATAACAACTAATT 987
 Db 1335 ILethrAspIleLeuAsnSerPheAsnThrPheSerPheThrAspLeuAsnIleIle 1354
 QY 988 GATGAGCGCTGTTTCTTCTATTCATCATTCGATTATCTCT-----TATCGATGATAT 1041
 Db 1355 AspAspProSerPheSerAspAspMetIleArgSerPheGlnLysLeuHisSerThrAsn 1374
 QY 1042 CATGAAAAAATCCAAACATGTTATCTTTACGAAACACTATTCAGAACCAACGTTAATC 1101
 Db 1375 TyrGlnAspIleLeuGlnItyrIleTyr----- 1383
 QY 1102 CCGACTTTTATAGATGATGATCATTTATTTCTCATATATAGATGATTAAGTATATT 1161
 Db 1384 -----GlnLeuAspAsnPheIleSer-----LysLysPheAsnLeuValSer 1397
 QY 1162 AACGATGATTTATTAATTAATAATTAATCAGCTGATTA-----CGTCAATGATGTA 1215
 Db 1398 LysLysAspIleValIleuPheGlnIleuGlnIleuLeuSerLysGlnSerLeuVal 1417
 QY 1216 AATACATATAT----- 1227
 Db 1418 SerPheAsnItyrProLeuHisPheLysSerSerLysLeuIleAsnProGlyItyrLeu 1437
 QY 1228 ---AATGATATGATTTATGCTTAATATCAATACA-----AATCATTAAGATGAT 1275
 Db 1438 GlnLeuHisGlnPheGlnIleSerAsnLeuPheThrIleSerThrIleu-----IleLeu 1456
 QY 1276 CAAGCATTCACGGGTATGATTCATGATCATATACGTTGTTATTCATATATGAA 1335
 Db 1457 LysAspAspAsnGlyThrGlnSerLeu----- 1465
 QY 1336 TGTGAATACTTTCATGACGCTGATTTATTTCAAACTATTTATTAACACACAGGT 1395
 Db 1466 -----PheGlnIleuLysLeuProGln 1472
 QY 1396 AAGTTAAAAACAATCAATATG---ACATCACCTTACGACTATCACTTACGTATGAT 1452
 Db 1473 SerIleLysLeuLeuIleLysLeuHisThrSerLeuThrPhePheValMetGlnIle 1492
 QY 1453 ATCAGCAGACCCATACATCAATATGAGGAGGTATGTTATCTTCACTGTTTAAATGGA 1512
 Db 1493 SerAsnValAsnLysSerSerSerGlnArgLeuThrCysLysValIleLeuGlnIle 1512
 QY 1513 TTATATGGCATACCTGCATTACGTTCACTTTTAACTATTCGTTTAGATATAACAAT 1572
 Db 1513 LeuAsnThrIleArgThrLysAsnGlySerLeuAspLeuPheAspSerGlnIleAspGln 1532
 QY 1573 GAACATATACATATC-----ATTACGGTTACAAAAACCT----- 1608
 Db 1533 SerProHisAlaIleCysProHisIleProAlaPheIleGlnThrAlaIleAlaHisAla 1552
 QY 1609 -----GAACGTATATATATATCTCTACATTTGTCACATCA----- 1644
 Db 1553 IleIleSerProLysSerArgAsnItyrGlnLeuSerThrIleLysAlaSerGlnLysLeu 1572
 QY 1645 -----CCTTCATTTGATAACTTATGCTCTTTC 1674
 Db 1573 SerAspProThrLysGlyThrGlnAsnLeuArgSerIleSerAsnValLeu----- 1589
 QY 1675 CATTACTTAACGAAAGTAATTAATGACAAATTTTAT-----TATGCAATACGTAT 1728
 Db 1590 -----GlnLysIleAspAspIleHisIleLysArgPheIleGlnIleAsp 1604
 QY 1729 AGTTGTATATGAATCC-----GTTGTTAAACCC 1758
 Db 1605 AspValPheSerLysAsnGlyLysAsnLeuCysProCysProGlyTyrPheIleSerArg 1624
 QY 1759 TTATTGAACCCAGCTTATTGACCCGATAGCCTTAGTAATGGATTTTGAACAAACGA 1818

Db 1625 LeuLeuGlnIleSerGlnPheValPro-----AsnMet 1635
 QY 1819 CAGATAGATAAATGTTGTTACTGATCATTAAGAAATTCGATATGAAGTGAAGAAC 1878
 Db 1636 SerIleThrAsnSerLysLeuIleAsnPheAspLysArgArgPhe---ValAsnAsnIle 1654
 QY 1879 ATTAATAATGCTTCTGCTGTTATCCGAAACCGCTTGTATACAGCGCTGATTTGAA 1938
 Db 1655 IleSerAsnValLeuAspLeuIlePro---AsnGlnIleGlnPheProLeuAspIleGln 1673
 QY 1939 -----ACCTTTGTACGTGAACAATTTCTTGACCGT 1968
 Db 1674 MetSerAspGlnAsnProSerLysArgThrPhe-----Gly 1686
 QY 1969 GCCATTTATGAACAAACAAAGATTCATATATGACGAAAGTACAAATATGATATACCG 2028
 Db 1687 ArgIleLeuPheAsnAsnPheGlnAspValAsnLys-----ValItyrArg 1701
 QY 2029 TCTAAACTGAAATGTATGCTGTAAT----- 2055
 Db 1702 LysLysThrLysLysValSerGlnIleAlaIleSerGlnArgPheGlnGlnGly 1721
 QY 2056 GATATGATGATATTTTACGATGACTAAT----- 2088
 Db 1722 ValPheAsnGlnIleLeuValAsnGlnIleGlnLysIleLysArgGlnAlaArgLysLeu 1741
 QY 2089 -----ATGAAACGTAATTTATTAAGACGCTAGAGAA----- 2124
 Db 1742 GlnValLeuLeuAspGlnGlnLysIleLeuLysAsnSerAlaAlaLeuHisGlnAlaVal 1761
 QY 2125 -----AATTCCACCACTAGTCAATTGATGATAT 2154
 Db 1762 ProLysLysAsnArgLysSerValIleIleSerGlyThrHisSerAspAsnAspHisSer 1781
 QY 2155 CTTTATATGAAAGACATCGTTCATTT---TCACCTTACGACTTATTTCCAGTTGAA 2211
 Db 1782 TyrAsnIleAsnLysAsnThrGlyGlnThrProSerLeuGlySerValMetGlnSerAsn 1801
 QY 2212 CGTTCATGATACAAATCTGAT 2235
 Db 1802 AsnSerAlaArgAsnArgArgAsp 1809

RESULT 3

US-09-774-414-3
 ; Sequence 3, Application US/09774414
 ; Patent No. US2002010231A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Institute of Physical and Chemical Research
 ; TITLE OF INVENTION: Endonuclease
 ; FILE REFERENCE: PH-651
 ; CURRENT APPLICATION NUMBER: US/09/774,414
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 09/306,970
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 476
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 ; US-09-774-414-3

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 156.50	476	116
Percent Similarity: 38.04%	Conservative: 97	
Best Local Similarity: 20.71%	Mismatches: 184	
Query Match: 3.85%	Indels: 164	
DB:	Gaps: 29	

US-09-727-892A-2 (1-2286) x US-09-774-414-3 (1-476)

QY 1798 -----AANTGGATATTTGAAAC----- 1815
Db 690 GlnArgValIysArgTPlYsAspSerTyrGlnTrpMetValSerAsnTrpLeuSerArg 709
QY 1815 ----- 1815
Db 710 IleThrGlnPheAsnHisIleAsnTyrGlnMetTyrAspSerLeuSerTyrGlnAla 729
QY 1816 GAACAGATACATAGATGTTTGACGATCATACAGAAATATGCATATGAGTGAATGA 1875
Db 730 AspAlaIleLysAlaLysIleAspLeuGluTyrLysTyrSerGlySerAspLysGlu 749
QY 1876 AAGATTAATTCCTCTCTGCTGGATACGAAACGCCCTTGATGATCAACGCCG----- 1929
Db 750 AsnIleLysSerGlnValGluAsnLeu---LysAsnSerLeuAspValLysIleSerGlu 768
QY 1930 -----GATTTGAAACCTTTGTACGTGACGACAAATTTCTTGACGGTGCCATTATTGA 1980
Db 769 AlaMetAsnAsnIleAsnLysPheIleArgGlu-CysSerValThrTyr---LeuPheLys 787
QY 1981 AACAATTAAGTATCTATATGAGCAAGTACATATCGATATATCCGTAAACTGAA 2040
Db 787 sAsnMetLeuProLysValIleAspGluLeuAsnLysPheAspLeuArgThrLysThrGln 807
QY 2041 ATTGTATGTGTATGTATGTATGATGATATTTACTGATG 2080
Db 807 uLeuIleAsnLeuIleAspSerHisAsnIleIleLeuVal 820

RESULT 5
US-09-801-368-354
; Sequence 354, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Shetman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 354
; LENGTH: 1427
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-354

Alignment Scores:
Pseq. No.: 1,19e-05 Length: 1427
Score: 152.50 Matches: 159
Percent Similarity: 34.24% Conservative: 133
Best Local Similarity: 19.14% Mismatches: 289
Query Match: 3.75% Indels: 291
DB: 10 Gaps: 44

US-09-727-892a-2 (1-2286) x US-09-801-368-354 (1-1427)

QY 88 GTTAACGACGAAACCAACCAATATATA-----AACCTACTATTCT 135
Db 344 IleAsnAspIleLysLysAsnAsnLysIleLysLeuAsnIleLeuLysIleLeuSer 363
QY 136 GTACGAATGGTGGTTGTTATGAAATGATGATGATGATTTCCG-----AGT 189
Db 364 LeuIleLeuLysIlePheGlnGlnSerLeuGlnValPheIlePheProThrSerAsn 383
QY 190 TTGCAATCTTTTATGACCAATTTTATACGATGCGAAAGCGATACATACAA 249
Db 384 TrpGluIleTyrLysProLeuLeuPheGlnIleValSerAsnAlaAspPheAsnGln 403
QY 250 TCAAAA----- 255
Db 404 SerAspMetLysLysLeuGlnLeuIleSerTyrArgAsnGlnSerLeuLysAsnAsn 423
QY 256 -----ACAGATATATCATGATTTGCATATACATGATATATAATAC----- 294
Db 424 SerSerIleArgAsnValIleMetSerAlaSerAsnAlaAspPheGlnLeuThrIle 443
QY 295 -----GATATCAT 303
Db 444 ValThrCysLysGlnPheProLysLeuSerCysIleGlnLeuAsnCysIleAspThrGln 463
QY 304 TTT-----TTACTTAAAGAC----- 318
Db 464 PheThrLysLeuLeuAspAspAsnProThrGluPheAspTrpProThrTyrValAspGln 483
QY 319 -----ACATG-----CGT 327
Db 484 AsnProLeuThrMetHisLysIleIleGlnLeuIleLeuTrpSerIleHisProSerArg 503
QY 328 TATTTGATATATTTACACGCCAAATATATATTTAA----- 366
Db 504 GlnPheAspHisTyrGlnSerAsnGlnLeuValAlaLysLeuLeuLeuArgIleAsn 523
QY 367 TCTGCAGAGAAATGACACACATTAAATGAAGAGCTACT----- 411
Db 524 SerThrAspGluAspLeuHisGlnPheGlnIleGluAspAlaIleTrpSerLeuValPhe 543
QY 412 ATTTAGCCAAATCAAAATGTAATTTAGAAAAAGCTTAATCTCATCATTTA 471
Db 544 GlnLeuAlaLysAsnPheSerAla-----GlnLysArgValValSerTyrMetPro 561
QY 472 GATTTACAAATGTTTAAATGTTTAAATTTAAATTTATGATATGATGAAACC 531
Db 562 SerLeuTyrArgLeuLeuAsn-----IleLeuIleThrTyrGlyIleIleLysVal 578
QY 532 AATACATCATTTGCAACATTAGTAAGAAATTTACTGATGGT-----TATTTACA 585
Db 579 ProThrTyrIle-----ArgLysLeuIleSerSerGlyLeuLeuTyrLeuGln 594
QY 586 GATCACAA----- 594
Db 595 AspSerAsnAspLysPheValHisValGlnLeuLeuIleAsnLeuLysIleSerProLeu 614
QY 595 CTTAAACAGATTTTAAATTTACGATTTTGAATAAGATTAATGATGATGTGAA 654
Db 615 MetLysSerGlnTyrAsnMetValLeu-----ArgAsn 625
QY 655 GCTATGACTATGCTGTAATGTTTGCAAACTGCACACTGCAACACTTACATACAT 714
Db 626 ValMetGluTyrAspValLysPheTyrGluIlePheAsnPheAspGlnLeuValGluIle 645
QY 715 CATATGACGTATATATATAGTATGTCATATTCATATTAATGATATATTTCCAA 774
Db 646 ThrGluGluIle-----LysMetArgIleLeuSerAsnAspIle---ThrAsn 660
QY 775 TTGACTATTAACAAATTAACATTTTCATGATATATTAG-----GAATCTACTGAT 828
Db 661 LeuGlnLeuSerLysThrProLeuSerIleLysIleMetValAlaGluTrpTyrLeuSer 680
QY 829 -----AATGAATGACAGCTTTTCACTACTACCAACCAATATGAA 867

[illegible]

OY	1486	-----	ATGTTATCTAAA	1497
Db	314	SeerGUThrThrllelyrAlaaspSerAargasnValGluaspGlyArgSerThlGlnSer	333	
OY	1498	GTCCGTTTAAATGGATATATATGGCATACCGTCATACGTCATTCATTTAACTTATTCGT	1557	
Db	334	ValleuMetSerAlaLeuaspGlyPheasnIleleArgTylGlnValPheThrPheYs	353	
OY	1558	TTAGATATATACATGACATGACATATAC-----AATATCATTAACGGTTACCAAAAC	1605	
Db	354	MetasnaspLysGlyGluAlaIleaspLysaspGlyasnLeuValThr-----ASP	370	
OY	1606	ACTAAGACATATATATATTATTC-----TCTACATTTGTGCATACGTCATTCGTAT	1656	
Db	371	SerSerLysleuValleuPheGlyLysaspPlysglutTThrGlyGluaspLysPhe	390	
OY	1657	AACATTATTTGTCCTTCTCCAACTACTTAACGGAAGGAATTCACGACATTTTATTTAT	1716	
Db	391	AsnVal-----GluAlaIleLysGlu-----AspLysSerleuLeu	403	
OY	1717	TGCCATACCTGATCTATTGTATATGAATACCGTGTGTAACCCCTTATTAACCCACGTTTA	1776	
Db	404	IleasphTrLysProValasnLeu-----SerMetaspLysAsnTylPheasnProSer---	421	
OY	1777	TTTCACCCGATAGCCTTAGTAAATGGGATATTGAAACAGACATAGATTAAGTCTTT	1836	
Db	422	-----LysSerAsnLysIleTyr	427	
OY	1837	GTACGATCATATAGAATATATGCATATGAAGTGAATGAAGATTAAATTTCTCTGCT	1896	
Db	428	ValArgasnProGluPheTyr-----LeuArgGlyLysIleSer---	440	
OY	1897	GGTATCCGAAAAAGCCCTTT-----	1917	
Db	441	-----AspLysGlyGlyPheasnTrpGluLeuArgValasnGlnSerValValaspasn	458	
OY	1918	-----GATACACGCGTCGATTTTGGAAACCTTTTACGT	1950	
Db	459	TyrlleuIleTyrGlyaspPheLhnsIleasphsThrArgaspPheasn-----IleYs	476	
OY	1951	GACACATCTTTGACGCGTCCCATTTATGAAAACAATAAAGATCTATATATAGCAAGT	2010	
Db	477	LeuAsnValLysaspGlyaspIleMetaspTrpGlyMetLysaspTyrLysAlaasnGly	496	
OY	2011	ACAATATGATATATCCGTCGTAAACGAATGTATGTGTCGATATGATATGATGAATAT	2070	
Db	497	-----PheProaspLysValThrAspMetaspLysasnValTyrleuGlnThr	512	
OY	2071	TTTACTGATGACTTAATATGAAACGT-----GAATTATATTAAGAAC	2115	
Db	513	GlyTyrSeraspLeuAsnAlaLysAlaValAlGlyValHisTyrGlnPheleuTyraspasn	532	
OY	2116	GCTAAGAAA-----AATTGCGACCATAGTCATTTGATGATGATCTTTATATTAAGT	2168	
Db	533	ValLysProGluValasnIleaspProLysGlyAsnThrSerleuTyrAlaaspGly	552	
OY	2170	GACATCGGTCATTTTCACCTTAAGAC	2196	
Db	553	LysSerValValPheasnIleasnasp	561	
RESULT 7				
US-09-815-242-11609				
Sequence 11609, Application US/09815242				
Patent No. US20020061569A1				
GENERAL INFORMATION:				
APPLICANT: Haseldeck, Robert				
APPLICANT: Ohlsen, Karl L.				
APPLICANT: Zyskind, Judith W.				
APPLICANT: Wall, Daniel				
APPLICANT: Trawick, John D.				
APPLICANT: Carr, Grant J.				
APPLICANT: Yamamoto, Robert T.				

RESULT 7
HC 00 01E 343 11600

; Sequence 11609, Application US/09815242

; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T

.....

```

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11609
LENGTH: 897
TYPE: PRF
ORGANISM: Helicobacter pylori
US-09-815-242-11609

Alignment Scores:
Pred. No.: 2.7e-05 Length: 897
Score: 148.00 Matches: 178
Percent Similarity: 33.45% Conservative: 121
Best Local Similarity: 19.91% Mismatches: 315
Query Match: 3.64% Indels: 280
DB: 10 Gaps: 46

US-09-727-892a-2 (1-2286) x US-09-815-242-11609 (1-897)
QY 67 GAACATTGACGTACATTAAGTAAAGCGAAAAACCAACCAATATAAAACGTT 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 GlnLysLeuGlyGluTyrLysGlnasn---ArgLysAspAlaProLysGluMetLeu 96

QY 127 ACTTATCTGTACCAATTCGTTGTTT-----AAT 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 GlnIleProIleLeuLysLeuTyrPheLysGlnLysMetGlyPheThrCysValGluValGly 116

QY 157 GGTATGAAATGATGTTGAAGTATTCGAGTTTCGATCTTTTATGACCATTTAT 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 GlyPheGluAlaasp---AspValIleAlaSerLeuAlaThrLeu---SerProTyrLys 134

QY 217 ACG-----TAGTGAAGAAGCGTGAT-----ACAATC 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 ThrArgIleTyrSerLysAspLysAspPheAsnGlnLeuLeuSerAspLysIleAlaLeu 154

QY 244 ACAAAATCAAAAACAGATTTATCATGATTCGACATTAAGT---AATTAATNC----- 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 PheAsnGlyLysThrGln-----PheLeuAlaLysAspCysValGluLysIleGlyIle 172

QY 295 -----GATAATCATTTTACTTAAGACACCATGCGTTATTTGAT 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 LeuProSerGlnPheThrAspLysGlnGlyIleValGlyAspSerAspAsnTyrLys 192

QY 337 AATATTACAGCGAAATATATTAATATCGAAGAAGAAATGATATTTAGAA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 GlyVal-----LysGlyIleGlySerLysAsnAlaLysGluLeuLeuGlnArgLeuGly 210

QY 394 -----AAATGAAGAGGCTACTATTATAGCCAAATAATGATATTTAGAA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 SerLeuGluLysIleTyrGlnAsnLeuAspLeuAlaLysAsn-----LeuLeuSer 227

QY 445 AAACGCGTTAAATCTTCAATCAATTTAGATTAAACAATGTTTAAATGCGTTTAAATTT 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 ProLysMetTyrGlnAlaLeuIleGlnAsp----- 237

```

```

QY 505 AATATTATGATTAACCTTATGAAAAACAATACATCAATTTGGACATTAAGTAAGAAATTA 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 -----LysGlySerAlaPheLeuSerLysGluLeu 247

QY 565 -----CTTGATGCGTGTATTTAAACAGATTCACACTTAAACAGATTTTAATATACG 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 AlaThrLeuGluArgGlyCysIleLysGlu-----PheAspPheLeuSerCys 263

QY 619 ATTTTGATTAAGATAAGATATG-----AATGATAGGAAGCGCTATGACAT 666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 AlaPheProSerGlnAsnProLeuLysIleLysAspGluLeuLysGluTyrGlyPhe 283

QY 667 GCTGGAATGTTTGGAAAACCTCACACCTGACACACTTACATCAATTAATGACGTG 726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 ---IleSerThrLeuArgAspLeu-----GluAsnSerProPheIleValGluAsnVal 300

QY 727 ATTTATTAAGTATGCGCATTCATTCATATAGTGATATTTCCAAATTTTGACATATAAC 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 ProIleLeuAsnSerThrProIle-----LeuAspAsnThrProAlaLeuAspAsnAla 318

QY 787 AATTAACATTTTCATTCGAATATTAATGAAATCT----- 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 ProLysLysSerArgMetIleValLeuGluSerAlaGluProLeuSerMetPheLeuGlu 338

QY 820 TACTTGATTAATGAATACACGT---TTTCAGTTACTCAACCAATATCAAGATATATAA 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 LysLeuGluAsnProAsnAlaArgValPheMetArgLeuValLeuAspLysAspLysLys 358

QY 877 ATATCTTATACATTAATTAATTCATTCAGATATGATTTTATGACTATTAATATCA--- 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 IleLeuAlaLeuAlaPheLeuLeuGlnAspGlnGlyTyrPheLeuProLeuGluGluAla 378

QY 934 ---TTCTATCGTGTGTTTAAATATGTATTAACCAACCAATTAACCAACTATTTGAT 990
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 LeuPheSerProPheSerLeuGluPheLeuGlnAsnAlaPhe---SerGlnMetLeuGln 397

QY 991 GAGCCTGTTTTCATTCATGACATCAATTCGAGTATTCCTTATGAGTTCATGAGAAA 1050
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 HisAlaCys-----IleIleGlyHisAspLeu 406

QY 1051 ATTCACATGTTATACCTTTTACGAACACTATTTCAGAACCA----- 1092
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 LysProLeuLeuSerPheLeuLysAlaLysTyrGlnValProLeuGluAsnIleArgIle 426

QY 1093 -----ACGTAATCCGACTTTTATGAGATGACACATTAATTTTCATATATTAAGAT 1146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 GlnAspThrGlnIleLeuAlaPheLeuLysAsnProGluLysValGlyPheAspGluVal 446

QY 1147 GATAAAGATGATTTTAACGATGATTTATTAATAAATTAATCAACGCTGATTAACGTCAA 1206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 LeuLysGluTyrLeuLysGluAspLeu-----IleProHisGluLysIleLysAsp 463

QY 1207 ATGATTTGTAATAATCTAATAAATGATAATGATTAAGTTAATATACATTAACATTA 1266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 PheLysThrLysSerLysAlaGluLysSerGluLeuLeuSerMetGluLeuAsnAlaLeu 483

QY 1267 AGAATGATTCAGACATTAACGATTAATGATTCATACATATACGTTGTTGTT 1326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 LysArgLeu----- 486

QY 1327 ATATATGAATGTCAATCTTCAT-----GCA 1353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 -----CysGluTyrPheGluLysGlyLysGluLeuGlnAspLeuThrLeuAla 503

QY 1354 CGTGATATTT----- 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 ArgAspIleGluThrProPheValLysValLeuMetGlyMetGluPheGlnGlyPheLys 523

QY 1363 ATTTTCAAAACGATTTTATTAATAACCAAGGTAAGTTAAAAAACAATCAATATGACAA 1422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 IleAspAlaProTyrPheLysArgLeuGlnGluPheLysAsnGlnLeuAsnVal--- 542

```


QY 1423 TCACCTTACGACATTCACATTACTGATGATATACACGACACCCATATCTCAATAGAGAG 1482
 Db 543 -----LeuGluArgGlnIleLeuAspLeuIleValAspPheAsnLeuAsnSerPro 560
 QY 1483 GTATTGTTAACTAGTTCGTTTAAATGATTTATATGATGACATACCTGATTACGTTACAT 1542
 Db 561 LysGlnLeuGluValIleLeuTyrAspLysLeu---GlyLeuProLysAsnLysSerHis 579
 QY 1543 TTT-----AACTATTCGTTTAGATGATATACAGTAACTATACATATCATTT 1590
 Db 580 SerThrAspGluLysAsnLeuLeuLysIleLeuAspLysHisProSerIleProLeuIle 599
 QY 1591 AACGGTACAAAACACTGACGATATATATTTCTCTCATTCATTTGACATGACATGACGTCA 1650
 Db 600 LeuGluTyrArg-----GluLeuAsnLysLeuPheAsnThrTyrTrpTrp----- 614
 QY 1651 TTGTAACTATTAGTTCCTCTTCCAACTACTTAACGGAAGTGAA-----ATTGACGAC 1704
 Db 615 -----ProLeuLeuArgLeuLysAspLysAspAspLysIleHisThr 628
 QY 1705 AATTATTATTATTCGATACGATAGTTTGATATGAAATCCGTTTAAACCTTTATG 1764
 Db 629 ThrHelIeGlnThrGlyThAlaThrGlyArgLeuSerSerHisSerProAsnLeuGln 648
 QY 1765 AACCCAGTTTATTCGACCCGATAGCCTTAGTAAATGGGATATTGAAAAGACAGATA 1824
 Db 649 AsnIleProValArgSerProLysGlyLeu-----LeuIle 660
 QY 1825 GATAGATGTTTGTTACTGATCATATAAGAAATATCATATGACATGCAAGATTAA 1884
 Db 661 ArgLysGlyPheIleAsnSerLysGluIleCysLeu----- 673
 QY 1885 ATTCCTCTGCTGATATACCAAAAGCCCTTGATCAACGCTGATTTGAACCTTT 1944
 Db 674 -----LeuGlyValAspTyrSerGlnIle 681
 QY 1945 -----GTACGTGAACAATTTCTTGAC 1965
 Db 682 GluLeuArgLeuLeuAlaHisPheSerGlnAspLysAspLeuMetGluAlaPheLeuLys 701
 QY 1966 GGTGCCATTATT-----GAAACATATAAGTATCTAATATAG 2004
 Db 702 GlyArgAspIleHisLeuGluThrSerLysAlaLeuPheGlyGluAspLeuAlaLysGlu 721
 QY 2004 ----- 2004
 Db 722 LysArgSerIleAlaLysSerIleAsnPheGlyLeuValTyrGlyMetGlySerLysLys 741
 QY 2005 ---CAAGGTACATATGATATATCCGCTAAACGTGAATTTGATGTGTAATGTATAT 2061
 Db 742 LeuSerGlnThrSerIleProLeuSerGluAlaLys-----SetTyrIle 757
 QY 2062 GATGAATATTTT-----ACTGATGACTTAATGAAACGTGAATTT 2103
 Db 758 GluAlaTyrPheLysArgPheProSerIleLysAspTyrLeuAsnGlyMetArgGluGlu 777
 QY 2104 ATATTAAAGACGCTAG-----GAAATTCGACCAATACT 2139
 Db 778 IleLeuLysThrSerLysAlaPheThrLeuLeuGlyArgTyrArgValAlaPheAspPheThr 797
 QY 2140 CAATTGATGATTTCTT-----TATATTGAAGTGAATCGGTTCTATT----- 2184
 Db 798 GlyValAsnAspTyrValLysGlyAsnTyrLeuArgGluGlyValAsnAlaIlePheGln 817
 QY 2185 ---TCACTTAACGACTA-----TTCCAGTTGAACGTTGATGACATTAAC 2226
 Db 818 GlySerAlaSerAspLeuLeuLysLeuGlyMetLeuLysValSerGluArgPheLysAsn 837
 QY 2227 AAATCTGATTGATATATTAACGTGAACGTGAATGATGAATA 2268
 Db 838 AsnProSerValArgLeuLeuLeuValHisAspGluLeu 851

US-09-843-676-8
 ; Sequence 8, Application US/09843676
 ; Patent No. US20020164786A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; ; Lindner, Joachim
 ; ; Nakamura, Toru
 ; ; Chapman, Karen B.
 ; ; Morin, Gregg B.
 ; ; Harley, Calvin
 ; ; Andrews, William H.
 ; TITLE OF INVENTION: No. US20020164786A1el Telomerase
 ; NUMBER OF SEQUENCES: 225
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/843,676
 ; FILING DATE: 26-Apr-2001
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/854,050
 ; FILING DATE: 09-MAY-1997
 ; APPLICATION NUMBER: US 08/846,017
 ; FILING DATE: 25-APR-1997
 ; APPLICATION NUMBER: US 08/844,419
 ; FILING DATE: 18-APR-1997
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 01-OCT-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Apple, Randolph T.
 ; REGISTRATION NUMBER: 36,429
 ; REFERENCE/DOCKET NUMBER: 015389--002930US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 872 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: No. US20020164786A1 Relevant
 ; TOPOLOGY: No. US20020164786A1 Relevant
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 ; US-09-843-676-8
 ;
 ; Alignment Scores:
 ; Pred. No.: 4.89e-05 Length: 872
 ; Score: 145.00 Matches: 132
 ; Percent Similarity: 37.31% Conservative: 134
 ; Best Local Similarity: 18.51% Mismatches: 247
 ; Query Match: 3.56% Indels: 200
 ; DB: 9 Gaps: 37
 ; US-09-727-892a-2 (1-2286) x US-09-843-676-8 (1-872)
 ; QY 79 TACATTAAGCTTAACGACGAGAAAAACCAACCAATATATAAAGCTTACTATTCTGTA 138
 ; Db 198 PheAsnLysAsnAsn-----TyrAspHisLeuAsnValSerIle 210
 ; QY 139 GCAATTGGTTGGTTTAATGTTGATGAATTTGATGATGATTTCCGAGTTTCGATCT 198
 ; Db 211 -----AsnArgLeuGluThrGluAlaGluPheTyrAla----- 221

QY	199	TTTTATGACCGATTTTATACGTAATGGGAAAGACGGTATACAACTACAAATCAAAACAA	258
Db	222	---PheaspaspPheSerGlnThrIleLys-----LeuthrsmksnSerTyr	236
QY	259	GATATTTATCATGATTTGGCAATTAACGTATTAATAATACGATTAATCATTTTCTACTTAAGAC	318
Db	237	GlnThrValaenIleasPValasn-----PheaspAsnAsnLeuSerIleLeuVala	253
QY	319	ACCATGGGTATTTTGGATATATTATTTACACGGCAAAATATAT-----TAT	360
Db	254	LeuLeuThrPheLeuLeuSerLeuGlnArgPheAsnIleLeuAsnIleArgSerSerTyr	273
QY	361	TTAAATCTGCAGAAAGAAATGAAACACACATTTAAATGAAGAGCGCTATTTTAGCC	420
Db	274	ThrArgAsnGlnTyrAsnPheGlnLysIleGlyLeuLeuGlnLr---ThrIlePheVala	292
QY	421	AAAAATCAAAATGTAATTTTNGAAAAACGGTGTAATCTTCATCATCAATTACATTTTACA	480
Db	293	-----ValValaPheSerHisArgHisLeuGlnGlyIleHisLeuValaPro	308
QY	481	ATGTTTTTAAATGGTTTTTAAATTTTAAATATTATTTGATTAACCTTATGAAAACCAATACACA	540
Db	309	-----CysGlnIlaPheGlnTyr-----LeuValaAsnSerSerSerGln	321
QY	541	ATTGCACAACTTAGGTAGAAGAAATTAATTCTGATGGTGGTTAATTACACAGATCACAACTTAA	600
Db	322	IleSer-----ValIlyAspSerGlnLeuGln	330
QY	601	ACAGATTTTATTAATACGATTTTGTATTAAGATAATAGATATGATAGATAGTACAAACCCAT	660
Db	331	Val---TyrSerPheSerThrAspIleLysValaAspThrAsnLysValaIlnAspTyr	349
QY	661	GACTATGCTGTGAATAGTTTGGCAAACTCAC-----CCTGAA	699
Db	350	PheLysPheLeuGlnGlnPheProArgLeuThrHisValSerGlnIlnIlaIleProVal	369
QY	700	CAACTTACA---TACATTCATTAAT---GACGTCATTAATTTAGGATATGCCAT-----	747
Db	370	SerAlaThrAsnAlaValGluAsnLeuAsnValLeuLeuLysValaLysHisIlaAsn	389
QY	748	-----ATTTCATTAAGATATATTTTCCAAATTTTAC	780
Db	390	LeuAsnLeuValSerIleProThrGlnPheAsnPheAspPheTyrPheValaAsnLeuGln	409
QY	781	TATACAAATTAACATTTTCATTCGTAATTAATTATTAAGAACTTACTTGAATATGAATGACA	840
Db	410	HisLeuLysLeuGlnPheGlyLeuGln-----ProAsnIleLeuThr	423
QY	841	CGTTTCAGTATACCAACCAATATCAAGATTAATAATCTTTATACACATATCATATTC	900
Db	424	LysGlnLysLeuGlnAsnLeuLeuSerIleLysGlnSer---LysAsnLeuLysPhe	442
QY	901	CATGATATGGAATTTTATGACATATTT-----AAATCATCTTATCGT	942
Db	443	LeuArgLeuAsnPheTyrIlnTyrValaIlnGlnIlnThrSerArgLysGlnIleLeuLys	462
QY	943	GGTGCTTAAATATGATATACACAAA-----	969
Db	463	GlnAlaThrThrIleLysAsnLeuLysAsnAsnLysAsnGlnGlnGlnIlnThrProGlnIlnThr	482
QY	970	-----TACATAACAAA	981
Db	483	LysAspIlnThrProSerGlnSerThrSerGlyMetLysPhePheAspHisLeuSerGln	502
QY	982	CTAATTATATAGCCTGTTTCTCTATTGCAATC-----AATTCG	1020
Db	503	LeuThrGlnIleuGlnLysPheSerValaLeuGlnAlaIlnThrGlnGlnIleTyrAspSer	522
QY	1021	AGTTATCCCTATGTGANGATCATGAAAAATCCCAACATCGTATTTACTTTTACGAAC	1080
Db	523	LeuHisLysLeuLeuIleArgSerThrAsnLeuLysLysPheLysLeuSerTyrIlnLysTyr	542
QY	1081	TATTCAGAACCAACGTTAAATCCCTACTTTTATAGATGATACAAATTAATTT-----	1131

Db 340 LeuTyrLysMetLeuThrGluIleTyrThrGluAspAsnPhenValLysPhePheLysVal 359
 QY 454 -----AAATCTCAATCAATTTAGATTAAACAATGTTTTAAAGTGTAAATTT 504
 Db 360 LeuAsnArgLysThrTyrLeuAsnPheAspLysAlaVal -----PheLysIle 375
 QY 505 AATATTAT -----GATACCTTTTGAAGAACCATACATCA 540
 Db 376 AsnIleValProLysValAsnTyrThrIleTyrAspGlyPheAsnLeuArgAsnThrAsn 395
 QY 541 ATTGCA ----- 546
 Db 396 LeuAlaIleAsnPheAsnGlyGlnAsnThrGluIleAsnAsnMetAsnPhenThrLysLeu 415
 QY 547 -----ACA 549
 Db 416 LysAsnPhenThrGlyLeuPheGluPheTyrLysLeuLysValArgGlyIleIleThr 435
 QY 550 TTAGTAAAGAATTACTGATGTTGTTAT -----TTAACAAGATCAACAATTAA 600
 Db 436 SerLysThrLysSerLeuAspLysGlyTyrAsnLysAlaLeuAsnAspLysIleLys 455
 QY 601 ACAGATTATTAATPACGATTTT -----GATAAAGATTAATGATG 642
 Db 456 ValAsn -----AsnTrpAspLeuPhePheSerProSerGluAspAsnPhenThrAsnAspLeu 474
 QY 643 AATGATAGTGAAGCCTATGACTGCTGTAAGTTTTCGAACATCACACCTGAAACA 702
 Db 475 AsnLysGlyGlu -----GluIleThrSerAspThr 484
 QY 703 CTACATACATCTCAATAGAGTATTATATTAGTATGTCATATTCATATTATAGAT 762
 Db 485 AsnIleGluAlaIleGluGlnAsnIleSerLeuAspLeuIleGlnGlnTyrLysLeuThr 504
 QY 763 ATATTCCAAATTTTGACTATACAAATTAACATTTTCATTTGAATATTATGAACTTAC 822
 Db 505 Phe -----AsnPheAspAsnGluProGluAsnIleSerIleGluAsnLeuSerSerAsp 522
 QY 823 TTGAATATGAAATG -----ACAGTTTTCAGTATACCAACCATAT 864
 Db 523 IleIleGlyGlnLeuGluLeuMetProAsnIleGluArgPheProAsnGlyLysLysTyr 542
 QY 865 CAAATATTTAAATATCTTATACACATTAATCATTTCCATGATGATGATTTTATGACTAT 924
 Db 543 GluLeuAspLys -----TyrThrMetPheHis -----TyrLeuArg 554
 QY 925 ATTAATCATCTCTATCGTGGTGTAAATATGATATAACACCAATATCATTAACAACATA 984
 Db 555 AlaGlnGluPheGlnIleHisGlyLysSerArgIleAlaLeuThrAsnSerValAsnGluAla 574
 QY 985 ATGATGAGCCTGT -----TTTTCTATGAC -----ATCAAT 1017
 Db 575 LeuLeuAsnProSerArgValTyrThrPhePheSerSerAspTyrValLysLysValAsn 594
 QY 1018 TCGAGTATCCTTATGATGATTCATCAGAAATAATCCACACAGCTATAC -----TTTAC 1074
 Db 595 LysAlaThrGluAlaIleMetPheLeuGlyTyrPvalGlnGluLeuValTyrAspPheThr 614
 QY 1075 GAACATATTCAGAACCAAGCTTAATCCACTTTTGAAGTATGATGATCAATATTTTCA 1134
 Db 615 AspLysThrSerGluValSerThrThrAspLysIleAlaAsp-IleThrIleIleIlePr 634
 QY 1135 TTATATAG -----ATTGATAAAGATGTTTAAAGCATATTA 1173
 Db 634 OTyrIleGlyProAlaLeuAsnIleGlyAsnMetLeuTyrLysAspAspPheValGlyAl 654
 QY 1174 -TTATATTAATTAATCAGCTGATATACGTAATGAT ----- 1212
 Db 654 AlaLeuIlePheSerGlyAlaValIleLeuLeuGluPheIleProGluIleAlaIleProVa 674
 QY 1213 -----GTAAATATCTATATTAAT ----- 1230
 Db 674 IleGlyThrPheAlaLeuValSerTyrIleAlaAsnLysValLeuThrValGlnThrIle 694
 QY 1231 -----GATATGATTAACGTTAATATCATAC 1256
 Db 694 eAspAsnAlaLeuSerLysArgAsnGlnLysTrpAspGluValTyrLysTyrIleValIh 714
 QY 1257 AATATCATTTAGAAATGATTCACAGACATTTACGGATATTGATTCATATACGTTTAA 1316
 Db 714 AsnTrpLeuAlaLysValAsn -----ThrGlnIleAspLeuIleArgLysLysMetLys 732
 QY 1317 TTGCTTTGTTATATATGAAGTAAATGATCTTCATGCACGATATATTTTCAAACTA 1376
 Db 732 SglAlaLeuGlnGlnAlaGlnIleAlaIleThrLysAla -----IleIleAsnTyrGlnTyr 750
 QY 1377 TTTTATTAATAACACAGATGATTAATAAACAATAATCATATGACATTCACCTTACAGTA 1436
 Db 750 AsnGlnTyrThrGlnGluGlu -----LysAsnAsnIleAsnPheAsn ----- 764
 QY 1437 TCACATTAAGTATGATATCAAGACACACCATCTCAATGAGAGAGTATGTTACTTA 1496
 Db 765 -----IleAspAspLeuSerSerLys -----LeuAsnGluSerIleAsnLysAlaIle 780
 QY 1497 AGTCGTTTAAATGATTA ----- 1515
 Db 780 tIleAsnIleAsnLysPheLeuAsnGlnLysSerValSerTyrLeuMetAsnSerMetIle 800
 QY 1516 -----TATGCAATA -----CCTGATTAACGTTACATTTTACTT 1550
 Db 800 eProTyrGlyValLysArgLeuGlnAspPheAspAlaSerLeuLysAspAlaLeuLeuLys 820
 QY 1551 ATTCCGTTTATGATGATTAACATGACATTAATCATATCATATCAAGTTTAAACAACCTGA 1610
 Db 820 sTyrIleTyrAspAsnArgGlyThrIleLeuIleGlyGlnValAlaPargLeuLysAspLysVa 840
 QY 1611 ACCTATATATTAATCTCTACATTTGTCACATCAGCTGATGATTAATGATTTGCTTCC 1670
 Db 840 IAsnAsnThrLeuSerThrAsp -----IlePr 849
 QY 1671 TTTCCAATACTTAACGAAAGTGAATGACAAATTTTATTTATGGATACGATG 1730
 Db 849 oPheGlnLeu -----SerLysTyrValAspAsnGlnArgLeuLeuSerThrPheThr 866
 QY 1731 TTGTATATGAATCCGTTGTTAAACCTTATTAACCCAGCTTATTAATGACCCGATAC 1790
 Db 866 rGluTyrIleLysAsnIleIle -----AsnThrSerIleLeu ----- 878
 QY 1791 CTTAGTAATGGGATATTAAGAACGACAGATAGATTAAGATTTTGTAOTGAATCATAA 1850
 Db 879 -----AsnLeuArgTyrGlnSerAsnHisLeuIleAspLeuSe 891
 QY 1851 GAAATATGATATGAAGTGAATGAAAGATTAATAATGCTTCTGCTGATACCGAATAA 1910
 Db 891 rArgTyrAlaSerLysIleAsn ----- 898
 QY 1911 CGCCTTTGATACAGGCTGATTTGAACCTTTGACGATACATTTCTTGACGCTGC 1970
 Db 899 -----IleGlySerLysValAsnPheAsp -----Br 907
 QY 1971 CATTAATGAACAAATAAAGTATCTATATGACAGATGACATATCATATATCCGTC 2030
 Db 907 oIleAspLysAsnGlnIleGlnLeuPheAsn -----LeuGluSerSe 921
 QY 2031 TAAACTGAATATGTAATGCTAATGATATATGATGATATTTTACATGAGACTTAATAT 2090
 Db 921 rLysIleGluValIle -----Le 927
 QY 2091 GAACGATGATTAATTAATAAGACCTAGAGAAATTCGACCATGATCAATTTGATGA 2150
 Db 927 uLysAsnAlaIleValTyrAsnSerMetTyrGlnAsnPhenSerThrSerPheTrp ----- 945
 QY 2151 TATCTTTATATTAAGAAAGTGCATCGTTCATTTGACCTTAACGACTTATTCGACGTGA 2210
 Db 946 -----IleArgIleProLysTyrPheAsnSerIleSerLeuAsnGlnLysTyrThrIleIle 964

QY 2211 ACCTTCAGTACATACCAATCT 2232
 Db 964 eaEncysmethyluasnasSer 971

RESULT 14

US-09-879-959-10
 ; Sequence 10, Application US/09879959
 ; Patent No. US20020160489A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Weigel, Paul H
 ; APPLICANT: Kumari, Kshama
 ; APPLICANT: Deangelis, Paul
 ; TITLE OF INVENTION: STREPTOCOCCUS EQUISMITIS HYALURONAN SYNTHASE GENE AND EXPRESSION
 ; FILE REFERENCE: IN BACILLUS SUBTILIS
 ; CURRENT APPLICATION NUMBER: US/09/879,959
 ; CURRENT FILING DATE: 2001-09-12
 ; PRIOR APPLICATION NUMBER: 09/469,200
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 09/178,851
 ; PRIOR FILING DATE: 1998-10-26
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 10
 ; LENGTH: 972
 ; TYPE: PRF
 ; ORGANISM: pasteurella multocida
 US-09-879-959-10

Alignment Scores:

Pred. No.: 9,08e-05 Length: 972
 Score: 142.00 Matches: 104
 Percent Similarity: 36.00% Conservative: 76
 Best Local Similarity: 20.80% Mismatches: 176
 Query Match: 3.49% Indels: 144
 DB: 9 Gaps: 25

US-09-727-892a-2 (1-2286) x US-09-879-959-10 (1-972)

QY 154 AATGGTATGAATTCATGTTGAGATTTCCGAGTTGCAATCTTTTATGACGATTT 213
 Db 569 AaGgLYTYTAsn-----TTPPGGLuPhe----- 576
 QY 214 TATACGTATGTGAAAAGAGCGTATACATCAACAAATCAAAAGATATATACATG 273
 Db 577 -----SerArgLulysLeuThrAlaMetIle 586
 QY 274 GCACATACCTGTAATAACGATATCATTTTCTTAAAGACACCATGCGTTATTTT 333
 Db 587 AlAhIs-----HisPheArgMet--PheThrIleArgAlaIrrp 598
 QY 334 GATAAATTTACACGCGAAATATATATTTAAATCGCAGAGAAATGAACACACATTA 393
 Db 599 HisLeuThrAspGlyPheSngIuIysIleGluAsnAlaValAspIrrpSphethelen 618
 QY 394 AAAATGAAGAGCGCTACTATTATTAGCCAAAATCAAAATGTAATTTTACAAAAAGCGTT 453
 Db 619 LysIleuSerGluValGlyLysPhe--LysHisLeuAsnLysIleCysTyIrrAsnArgVal 637
 QY 454 AATCTTCATCATTTAGATTTAACAATGTTTAAATGTTTAAATTTAATATTATT 513
 Db 638 -----LeuHisGly----- 640
 QY 514 GATACTTTATGAAACCAATACATCATTTGCAACATTAGGTAGAAATTAAGTATGTT 573
 Db 641 -----AspAsnThrSerIleLysLysLeuGlyIleGluLysLysAsnHis 655
 QY 574 GGTATTATACAGATCACACACTAA-----ACAGATTTTATATATAGAT 621
 Db 656 PheValValValAsnGlnSerLeuAsnArgGlnGlyIleThrTyIrrAsnTyIrrAspGlu 675
 QY 622 TTTCATTAAGATATGATATGATATGATATGATGACCTATGACATATGCTGTAAGTTT 681
 Db 622 TTTTATTAAGATATGATATGATATGATATGATGACCTATGACATATGCTGTAAGTTT 681

Db 676 PheAsp-----AspLeuAspGluSerArgLysTyr-----IlePhe 687
 QY 682 GCAAAACTCACA-----CTGAAACACTTCATACATTCATATATGACGTATATATTA 735
 Db 688 AsnIrrpThrAlaGluTyIrrGlnGluGluIleAspIleLeuLysAspIleLysIleIleGln 707
 QY 736 GGTATG--TGCATATTCATTTATAGCATATATTTCCAAAT----- 774
 Db 708 AsnLysAspAlaLysIleAlaValSerIlePheTyIrrProAsnThrLeuAsnGlyLeuVal 727
 QY 775 -----TTGACTATATACAAATTAACATTT-----TCATGATATAT 810
 Db 728 LysLysLeuAsnAsnIleIleGluTyIrrAsnLysAsnIlePheValIleValLeuHisVal 747
 QY 811 ATGCAATCTTACTGATATATGAAATGACACGTTTTCAGTTACTCAACAAATATCAAGAT 870
 Db 748 AspLysAsnHisLeuThrProAspIleLysLys--GluIleLeu----- 761
 QY 871 ATTAATAATCTTATACCATTCATTCATTCGATATGATATTTTATGACATATATTA 930
 Db 762 -----AlaPheTyIrrHisLysHisGlnValAsnIle----- 771
 QY 931 TCATTCATCGTGGTGGTTTAAATATGATATGATACACAAATATCAATTAACAAATATGAT 990
 Db 772 --LeuLeuAsnAsnAspIleSerTyIrrThrSer-----AsnArgLeuIleLys 787
 QY 991 GAGCCTGTTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1050
 Db 788 ThrGluAlaHisLeuSerAsnIleAsnLysLeuSerGlnLeuAsnLysCysGlu-- 806
 QY 1051 ATTCACACATGGTATATCTTTTACGACACATTCATTCACACACATTCATTCATTCAT 1110
 Db 807 -----TyIleIlePheAspAsnHisAspSer-----LeuPhe 817
 QY 1111 TTACATGATGACATATTTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1158
 Db 818 ValLysAsnAspSerTyIrrAlaTyIrrMetLysLysTyIrrAspValGlyMetAsnPheSerAla 837
 QY 1159 TTTACGATGATTTATTTATTAATAATTAATACACGCTATATTCGTCACAAATGATGTAA 1218
 Db 838 LeuThrHisAspTrpIleGluLysIleAsnAlaHisPropIrrpLysLysLeuIleLys 857
 QY 1219 TACTATATAATATGATATATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1260
 Db 858 ThrTyIrrPheAsnAspAsnAspLeuLysSerMetAsnValLysGlyAlaSerGlnGlyMet 877
 QY 1261 -----ACATTAAGATGATTCACAGACATTCGCGGTATTCATTCATTCATTCATTCAT 1314
 Db 878 PheMetThrTyIrrAlaLeuAlaHisGluLeu-----LeuThrIle 890
 QY 1315 AATTCGTTTGTATATATGATGAA-----TACTTCATTCATTCATTCATTCATTCAT 1356
 Db 891 IleLysGluValIleThrSerCysGlnSerIleAspSerValProGluTyIrrAsnThrGlu 910
 QY 1357 GATATATATTTTCAAAATCATTTTATTAACAAAGSTAGTAAACAAATCAAT 1416
 Db 911 AspIleThrPheGlnPheAlaLeuLeuIleLeuLysLysTrpGlyHisValPheAsn 930
 QY 1417 ATGACATACCTTACGATAT-----CACATTCATTCATTCATTCATTCATTCATTCAT 1455
 Db 931 LysThrSerThrLeuThrTyIrrMetProTrpGluArgLysLeuGlnTrpThrAsnGluGln 950
 QY 1456 AACGACACCATCTCAATAGAGAGGTATGTATTCATTAAGTCGTTTAAATGATTA 1515
 Db 951 IleGluSerAlaLysArgGlyLysAsnIleProValAsnLysPheIleIleAsnSerIle 970

RESULT 15

US-09-815-242-5251
 ; Sequence 5251, Application US/09815242
 ; Patent No. US2002061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.

```

: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: us/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ. ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5251
: LENGTH: 996
: TYPE: PRI
: ORGANISM: Staphylococcus aureus
US-09-815-242-5251

Alignment Scores:
Pred. No.: 0.000302 Length: 996
Score: 136.00 Matches: 142
Percent Similarity: 37.28% Conservative: 129
Best Local Similarity: 19.53% Mismatches: 290
Query Match: 3.34% Indels: 166
DB: 10 Gaps: 38

US-09-727-892a-2 (1-2286) x US-09-815-242-5251 (1-996)
QY 298 AATCATTTT-----TTACTTAAAGACACATGGTATTTTGATATATATACGCGAA 351
|||:|||||
Db 9 Asnashnheglp-rophleuylsugluilleps---pheserlyslleaspsasn 27
|||:|||||
QY 352 AATATATATTTAAATCTGCAGAGAAGAAATGACACACATTAAATAAGAGAGCTACT 411
|||:|||||
Db 28 Glueuheuileuieserlglysthglyserglysthrmetillephasalmet 47
|||:|||||
QY 412 ATTATCCCAAAATCAAAATGTAATTTAGAAAACGTGTTAAATCTTAATCAATTTA 471
|||:|||||
Db 48 Thrtyrzlaueuhhegllyslaserthrgluinargluuasnaspheuryg--- 66
|||:|||||
QY 472 GATTTAAATGTTTTTAATGCT-----TTTAATTTAATATT 510
|||:|||||
Db 67 -----SerhspshealaaspglylsglnPrometserValThrpheglupheglnleu 84
|||:|||||
QY 511 ATTGATTAACCTTTAGAAACCAATACATCAATTCGACACATTAGTAGAATTAAGTAT 570
|||:|||||
Db 85 Asnhisarglletrylrysvaahisarg-----Gln 94
|||:|||||
QY 571 GGTEGTTATTAAACAGAA---TCACAACTTAAACAGATTTAATATAGATTTTGGAT 627
|||:|||||
Db 95 GlyProtyrilleuylsuglyasnthrtThrglysthrasnalaalyspheaspaalphe--- 113
|||:|||||
QY 628 AAGATATATGATATGATAGTAGAGCCTATGACATGCTGTGAATGTTTGGAAA 687
|||:|||||
Db 114 -----Gluwetvalaspglylgs-----Tyrgluilear9gluserlysva 127
|||:|||||
QY 688 CTCACACCTGAACAACTTACATCATTCATTAATGACGTGATTATATAGTATGCGAT 747
|||:|||||
```

```

Db 128 lleserglythrglnpheille-----llegluileuenglyvalasna 142
QY 748 ATTCATTATGATATATTT-----CCAAT----- 774
|||:|||||
Db 143 AspGlnphearglnleuuhelleuProGlnGlylphelysargpheuileser 162
|||:|||||
QY 775 -----TTTGACTATACAAATTAACATTT 798
|||:|||||
Db 163 AsnserargluylsnglylleuargThrleuPhaspsersgluspheguala 182
|||:|||||
QY 799 TCATTTGATATATATGGAATCTTCTGAAATATGAATACAGCTTTGATACCAAC 858
|||:|||||
Db 183 llsargglulleuylsugluvalylslyslu-----Lyslaclnllleguasn 200
|||:|||||
QY 859 CAATATCAAGATATTAATAATCTTATACACATTAATTCATGATGATGATTTTAT 918
|||:|||||
Db 201 ArgtyrglnglnlleaspleuLeutrpglngluillegluserpheasp-----Asp 217
|||:|||||
QY 919 GACTATATTTAAATCATCTATCTGCTGTTTAAATGTATTAACCCAAATACATAAC 978
|||:|||||
Db 218 Asplysllelys-----GlyleuLeuvalalathrglnglnlleasp 232
|||:|||||
QY 979 AAACATATGATGAG---CCTTGTTTCTATGACATCAATTCAGATTATCCTATGTG 1035
|||:|||||
Db 233 LysvalillegluasnilleProleu-----LeuGlnlaIargserlysgluilleuala 250
|||:|||||
QY 1036 ATGTATCATGMAAAATTCACACATGTTATATCTTTACGACACTATTCAGAACACAG 1095
|||:|||||
Db 251 PhevalasngluserlysgluThrIalIelysgluThrIleleugluylslythr 270
|||:|||||
QY 1096 TTATGCCCTACTTTTACATGATGAC---AATATATTTCAATATATAGATGAT--- 1149
|||:|||||
Db 271 LeuGlnasnsnilleuylsaspasnlleasGlnLeuasnlystsnlyslleasphe 290
|||:|||||
QY 1150 -----AAAGATGATATTTAACGATGATTTATATTAATAATCAAGGTATTAT 1200
|||:|||||
Db 291 ValGlnleuylsuglnglnProgluIleugluilleglualalysleuylsleu 310
|||:|||||
QY 1201 CGTCAATG-----ATTGTAATATCTATATATATGATTAATGATTACGTTATATCAT 1254
|||:|||||
Db 311 GlnaspIlethrasnleuLeuasnTyrlleGuasnarglu-----LyslleGlu 327
|||:|||||
QY 1255 ACAATATACATTAAGATGATTCAGACATTCAGCGATTTGATTCATATACGATGTT 1314
|||:|||||
Db 328 ThrtlylIlealalysserlyslsaspIleeserGluThrasn----- 341
|||:|||||
QY 1315 AATTCGTTGTTTATATGAATGTAATCTTATCATGCACGCGATTTATTTTTCAAAAC 1374
|||:|||||
Db 342 AsnlyslleuasnleuaspCysasp-----LysargasnilleaspIlysglyl 358
|||:|||||
QY 1375 TATTTATTAACACACAGATTAAGTTA---AAAAAATTAATATGACATCACCTTAC 1431
|||:|||||
Db 359 LysmetleuGlnuasnnglyaspheullegluserlylleserhellleasplystr 378
|||:|||||
QY 1432 GACTATACATTTACTGATGATATCAAGACACACCATCTCA----- 1473
|||:|||||
Db 379 ArgVal---LeuPhasnspIleasnlystYrglnglnIserTyrleuasnilleguarg 397
|||:|||||
QY 1474 -----AATGAGGAGGTATGTTATCTAAGTCGTTTAAATGATATATAGCATACCT 1527
|||:|||||
Db 398 LeuargThrGlnaspIleuGlnleuGlyaspGluLeuasnasnleulleysgllyleuGlu 417
|||:|||||
QY 1528 GCATTACGTTACATTTTAACTTATTCGTTAGATGAT-----AAC 1569
|||:|||||
Db 418 LysValGlnaspsertlleGlyasnasnGlnuserAsptyGluylserlleleGluLeuasn 437
|||:|||||
QY 1570 AATGACATATACATTTCTATTAACGTTACAAAACACAGACAGTATATATATCTCT 1629
|||:|||||
Db 438 AsnAlaIlethrAsnIle-----AsnargGlnleuasnvalilleysglu 452
|||:|||||
QY 1630 ACATTTGTACATACAGCTTATGTTATTAATGATTCCTTCCATATATTAACGAA 1689
|||:|||||
Db 453 AsnGlnlyAlalysaspGluLeuaspIlyslleuGlyserlysglnglnleu---Glu 471
|||:|||||
```

```
QY 1690 AGTGAATTGACGACATTTTATTATTCGATGATAGTTGTATATGAAATCCGTT 1749
      :|||:|||||:
Db 472 AsnGlnIleAsnGlnGlu-----LysThrIle 480
      :|||:|||||:
QY 1750 GTTAACCCCTTATGTAACCCAGTTTATTCGACCCGATGCTTAGTAATGGATATT 1809
      :|||:|||||:
Db 481 LeuLysAsnLeu-----GluIleLysLeuAspArgTyrAsp--- 492
      :|||:|||||:
QY 1810 GAAACGACAGATAGATAGATGTTTGTATGATCATAGAAA--TATGCATATGAA 1866
      :|||:|||||:
Db 493 ---LysSerLysLeuAsp-----LeuAsnAspLysGlnSerPheIleSerGln 507
      :|||:|||||:
QY 1867 GTGATGCAAGATTAATTTGCTCT-----GCTGTATACCGAAAAACGCC 1914
      :|||:|||||:
Db 508 IleLysSerAlaValLysIleGlnLysGlnCysProIleCysGlnLysGlnIleGlnAsp 527
      :|||:|||||:
QY 1915 TTGTATACACCGCTGATTTGAAACCTTTGTACGTGACGACAAATCTTGCAGCTGCAT 1974
      :|||:|||||:
Db 528 LeuGlnLysIleHisLysPheAspSerIleAlaLysArgGln-----AsnGlnIle 544
      :|||:|||||:
QY 1975 ATGGAACAAATTAAGTATCTATATGACGCAAGGTACATATCGATATATCCGTATA 2034
      :|||:|||||:
Db 545 LysGlnIleGlnAlaAsnIleHisThrMetGlnSerAsnIleAlaValHisAsnSerGln 564
      :|||:|||||:
QY 2035 ACTGAATTTGTAATGTGTATGATGATGATGATATTTACTGATGAACTTAATAGAA 2094
      :|||:|||||:
Db 565 IleLysPheValAsnGlnLysIleSerAsnIleAsnIleLysThrGlnSerAspLeuSer 584
      :|||:|||||:
QY 2095 CGTGAATTATATTAAGAAGCGCTAGAGAAAATTTCGACCATAGTCATTTGATGATATT 2154
      :|||:|||||:
Db 585 LeuGlnValIleLysAsnLysArgLeuLeuGlnAsn-----GlnAsnAla 598
      :|||:|||||:
QY 2155 CTTTATATGAAAGTGACATCGGTTTCATTTTCACCTTAACGACTTATTTCCAGTTGACGT 2214
      :|||:|||||:
Db 599 LeuAsnAsnGlnArgAspLeuAsnLysPhe-----IleGlnGln 611
      :|||:|||||:
QY 2215 TCAGTACATACAAATCTGATTTGCATATA-----TTAAA 2250
      :|||:|||||:
Db 612 MetLysGlnGlnLysAspAsnLeuThrLeuGlnIleHisAsnLysGlnLeuArgLeuAsn 631
      :|||:|||||:
QY 2251 CGTGAACATGATGAATATAA 2271
      :|||:|||||:
Db 632 LysAsnGlnSerGlnLeuLys 638
      :|||:|||||:
```

Search completed: January 8, 2003, 19:33:59
Job time : 55.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 38 Seconds

(without alignments)
11566.489 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 4070

Sequence: 1 atgggtactactagaatgcac.....taaaaaaggaactgttaa 2286

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=ylh
-O=/cgs2.1/USPTO.spool/US09727892/runat.06012003.151201.9316/app_query.fasta.1.2439
-DB=PIR_73 -QMT=fastan -SUFFIX=rfp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl -DIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFM=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09727892 @CGN.1.1.39 @runat.06012003.151201.9316 -NCPU=6 -ICPU=3
-N-XY=XY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	257.5	6.3	960	2	S72284 DNA-directed RNA p
2	242	5.9	1121	2	F71613 hypothetical prote
3	241.5	5.9	575	1	ERBP29 DNA-directed RNA p
4	236.5	5.8	3973	2	B71612 hypothetical prote
5	234.5	5.8	372	2	UQ0161 DNA-directed RNA p
6	230	5.7	1802	2	G71616 hypothetical prote
7	229.5	5.6	1182	2	G71607 Probable integral
8	228.5	5.6	572	1	ERBP22 DNA-directed RNA p
9	225.5	5.5	568	2	S51275 DNA polymerase - p
10	221.5	5.4	1021	2	S26985 Probable DNA-direc
11	218	5.4	1817	2	D71606 hypothetical prote
12	215.5	5.3	3844	2	T18402 asparagine/asparta
13	213.5	5.2	1019	2	E90097 hypothetical prote
14	213	5.2	1306	2	T28313 ORF MSV152 probabl

15	212	5.2	807	2	T28279 ORF MSV119 probabl
16	211	5.2	1417	2	T18418 hypothetical prote
17	208.5	5.1	1162	2	I40817 botulinum toxin no
18	208	5.1	797	2	S28103 probable DNA-direc
19	208	5.1	1272	2	H82926 conserved hypothet
20	208	5.1	1817	2	H71611 probable secreted
21	207.5	5.1	547	2	S41618 probable DNA-direc
22	207.5	5.1	753	2	H90124 hypothetical prote
23	207.5	5.1	2136	2	A05037 hypothetical prote
24	207	5.1	1024	1	RNZ0BF DNA-directed RNA p
25	206.5	5.1	1532	2	T18438 hypothetical prote
26	206	5.1	1411	2	T18437 hypothetical prote
27	205.5	5.0	2437	2	T18482 hypothetical prote
28	205.5	5.0	4881	2	T18489 hypothetical prote
29	205	5.0	1188	2	A71621 protein with 5'-3'
30	202.5	5.0	2485	1	H71621 serine/threonine-s
31	201.5	5.0	2206	2	G71611 ORF MSV140 hypoth
32	200	4.9	608	2	T28301 ORF MSV140 hypoth
33	200	4.9	970	2	G84939 hypothetical prote
34	198	4.9	797	2	D71621 hypothetical prote
35	198	4.9	1162	2	A47708 hypothetical prote
36	197.5	4.9	1079	2	T28197 progenitor toxin n
37	197.5	4.9	1613	2	D90129 probable DNA-direc
38	196	4.8	2013	2	C71610 hypothetical prote
39	195.5	4.8	1465	2	A70199 probable membrane
40	195	4.8	1070	2	F90106 hypothetical prote
41	194.5	4.8	765	2	S72278 IAP100 protein [m
42	193	4.7	1008	2	T18508 ATP-dependent Clp
43	193	4.7	1516	2	E71619 RAD2 endonuclease
44	191.5	4.7	1247	2	E71616 hypothetical prote
45	191.5	4.7	1346	2	G71613 hypothetical prote

ALIGNMENTS

RESULT 1
S72284
DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain - Plasmodium falciparum plastid
N:Alternate names: RNA polymerase rpoC2
C:Species: plastid Plasmodium falciparum
C:Date: 14-Apr-1998 #sequence-revision 24-Apr-1998 #ext-change 20-Jun-2000
C:Accession: S72284
R:Wilson, R.O.M.; Denny, P.W.; Preiser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; J. Mol. Biol. 261, 155-172, 1996
A:Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum
A:Reference number: S72277; MUID:96346169; PMID:8757284
A:Accession: S72284
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-960 <WILD>
A:Cross-references: EMBL:X95275; NID:g1171583; PIDN:CA64574.1; PID:g1171589
A:Note: this apparently degenerate plastid is referred to as the apicoplast
A:Note: biosynthesis of this protein involves a -1 frameshift in the codon for resid
C:Genetics:
A:Gene: rpoC2
A:Genome: plastid
A:Note: DNA-directed RNA polymerase beta chain
C:Keywords: nucleotidyltransferase; plastid; transcription

Alignment Scores:
Pred. No.: 2.37e+08
Score: 257.50 Length: 960
Percent Similarity: 39.68% Matches: 189
Best Local Similarity: 23.08% Conservative: 136
Query Match: 6.33% Mismatches: 298
DB: Indels: 197
Gaps: 46

US-09-727-892A-2 (1-2286) x S72284 (1-960)
QY 126 TACTATTCTGTAGCAATGGTGTGTTAAGTTATGAATGATGATGATGATTC 185
DB 41 TysSerPheLeuYrAsnYrSerLeu-----AsnLeuYs 52

QY	186	GAGTGTGAATCTTTTATATGACCATTTTATAGCTATGCAAAAGCGATGATACATC	24
Db	53	AspHisSerAsnPhle-----TyrLeuLeuIleLeuTyrLysAsnLysIleAsn	70
QY	246	AAATTCAAAAACAGATATATATCATGATGTCACATAC--TGTAAATAATACATATCA	302
Db	71	IleTyrAsnAsnLysTyrTyrGlu-IleLysAsnAsnTyrIleAsnValPheLeuAsn	90
QY	303	TTTTTTTACTTAAGACCCGATGGGTATTTTGATATATATACCGCAAAATATAT--	360
Db	90	nTyrTyrTyrLeuLysValIleAsnLysIleGlnIleLeuAsnAsnAsnLeuTyrAs	110
QY	361	----TTAAATCTGCAGACAAAT-----GACACACATTTAAAAATGAA	401
Db	110	nLysIleAsnProIleTyrSerAsnLeuPheLeuPhePheAsnAsnLysIleLysIle	130
QY	402	AGAGCGATCATTT-----TTAGCGCAAAAAATCAAAATGT	434
Db	130	STYSerClnLeuGlnGlnLeuIleGlyTyrLysGlyTyrIleSerAsnIleLysGlyMe	150
QY	435	AATTTTGAAGAAAACGTGTTAAATCTTCACATCAATTTAGATTTAACAAATGTT--	486
Db	150	fileTyrClnLysProValIleAsnAsnTyrIleAsnGlnLeuAsnIleTyrGluTyrI	170
QY	487	-----TTAAATGGTTTAAATTAATATATATATGATCACTTAAAGAAACCAATACATC	539
Db	170	eleuSerCysTyrClySerLysLysGlyIleIleAspThrAlaLeuLysThr-----	187
QY	540	AATTGCAACATTTAGATGAGAATTTACTTGATGGTGGTTATTTTAAACAGATCAACAATTAA	599
Db	188	-----AlaspsSerGlyTyrLeuThrLysArgLeuIleAs	199
QY	600	AACGATTTTATATATACGATTTTGTATTAAGATATGATGATGAAGATAGAGAGCTTA	659
Db	199	nIleThrSerAsnPhleIleLysGlu-----	208
QY	660	TGACTATGCTGTGAATGTTTGTCCAAAACCTCACACCTGACAACTTACATACATTCATTA	719
Db	209	-----LeuAsnCys-----LysSerProPheIleLeuLysTyrIleLeuAs	222
QY	720	TGAGGTATATATATAGTATGCGATATGCAATTCATTTATAGTAAATATTTCCAAAT--	774
Db	222	n-----MetAspIleTyrGlyAsnIleI	230
QY	775	TTTGACTATACAAATTAACATTTTCATTTGAATATTANGAATCTTACTTGAATAATGA	833
Db	230	eleuProLeuAsnIleLeuArgPheLysIle--LeuGlnAsnAsnIleLeuAsn-LeuA	249
QY	834	AATGACAGGTTTACGTAGTATCAACCAATATCAAGATATAAATATCTTATATACATTA	893
Db	249	snaAsnGlyThrPheIleTyrThrLysAsnThrTyrIleThrLysTyrIleLeuAsnLysL	269
QY	894	TCATTTCCAGAT-----ATGAATTTTATGACATATATTAATCATCTTATCGTGGT	946
Db	269	euleuAsnLeuTyrAsnArgArgAsnIleLysIleuAsnIleLysSerValTyrLeu---	287
QY	947	GTTTAAATATGATATACCAACCAATATACATAACCAACTAATATATAGCGCTGTTTTTCTA	1006
Db	288	-----CysAsnIleTyrAsnAsn-----LleCysAsnThrCys-----L	299
QY	1007	TTTGACATCATTTGCAGTATACCTTA--GIGAGTATCATG	1045
Db	299	euaAsnTyrLysGlnLeuTyrLysTyrAsnLeuGlyInHisIleGlyValIleSerSerg	319
QY	1046	AAAAAATTCACACATGCTATACCTTTTACGAACACATCTACAGAACCAACG--	1096
Db	319	IuAlaIle-----SergIuProSerThrGlnMetV	329
QY	1097	TAACTCCTACTTTTATGATGATGACAAATATTTTCATATATATATAGATGATGAATGATG	1156
Db	329	AlleuArgThrPhe-----HisLaserSerIleLeuAsn--AspLysPheA	344

[illegible]

```

Db 689 yrasnlyrllasnasn-asniletyrasnleuynleuasnaspiletherlleglyleu 708
      ::::: ||||| ||||| ||||| |||||
QY 2120 GAGAAATTCGACCATAT-----GTCAATTCGATGATATTCCTTATATG 2164
      ::::: ||||| ||||| ||||| |||||
Db 709 Glnserlleasnilleephegluasnllysasnasnillepnephelieser 728
QY 2165 AAAGTGACATCGTTCATTTTCACCTATACGACTTATTCACAGTTG---AACGTTACGTAC 2221
      ||||| ||||| ||||| ||||| ||||| |||||
Db 729 Asnasnilletyrvalillepethyrllelystyrtyrasnlyrleuasnalleiletyr 748
QY 2222 ATACCAATCTGATTCGATATATTAACGTCGATGATGAAATGAAA 2272
      ||||| ||||| ||||| ||||| ||||| |||||
Db 749 lleyrasmn---llecyasnlystyrsnilleasnhslyrlystyrllys 764

RESULT 2
F71613
hypoethelial protein PFB0495w - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C/Accession: F71613
R/Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, H.O.
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MID:99021743; PMID:9804551
A/Accession: F71613
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1121 <GAR>
A/Cross-references: GB:AE001368; GB:AE001362; NID:93845197; PIDN:AACT8188.1; PID:9384519
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PFB0495w

Alignment Scores:
Pred. No.: 2,06e-07 Length: 1121
Score: 242.00 Matches: 201
Percent Similarity: 37.99% Conservative: 136
Best Local Similarity: 22.66% Mismatches: 291
Query Match: 5.95% Indels: 261
DB: 2 Gaps: 52

US-09-727-892a-2 (1-2286) x F71613 (1-1121)
QY 100 AAAAAACCAACCAATATAAAGCTTACTTATTCGTAGCATTTGGTTTAAATGTT 159
      ||||| ||||| ||||| |||||
Db 73 lysasnleuprollystyrllyscysalalystyrlclucyslleaserla----- 88
QY 160 TATGAATTCATGTTGAAGTATTCGAGTTGGATCTTTTATGACGCAATTTATACG 219
      ||||| ||||| ||||| |||||
Db 89 -----LysgluValIyrlYstyrlleuasnspoluyrlystyrcyspneasn 104
      ||||| ||||| ||||| |||||
QY 220 TAT-----GTGAAAAAGAGCGATGATCAATCAACAATACTA 252
      ||||| ||||| ||||| |||||
Db 105 TyrlleaserleucysaspilleeglnserVallysllephesapglueuasnplysTlr 124
QY 253 AAAACAGATATTCATGATTCGACATTAAGTAAATATGCAATATCAATTTTACTT 312
      ||||| ||||| ||||| |||||
Db 125 PheThrAspTyrAsnPhetyrllegluVallylasn---lleasplysasnValleuasn 143
QY 313 AAAGACCCATGCGTTATTTTGTATATAT-----TACACGGCAAAATATATA 359
      ||||| ||||| ||||| |||||
Db 144 LysIleasnluIleTyrlPheIlysasnllysasplleThrPheHlsArGagluIleleu 163
QY 360 TTATAATCTGCAGAGAAATATGACACATTAATAATGAAGAGGC----- 407
      ||||| ||||| ||||| |||||
Db 164 GlYlysllecyasnlysllemetserTyrlleHlsgluIleasnngluIleleu 183
QY 408 -----TACTATTTTACCCAAATAATGTAATTTTGAAGAAACGTTTAAATC 458
      ||||| ||||| ||||| |||||
Db 184 HlsPheleuIleTyrlPhePheArGtrPasnlysasnasn-----LysasnleuIleleu 201

```

```

QY 459 TTCAATCATTTAGATTTAACATGTTTAAATGTT---TTTAAATTTAAT----- 507
      ||||| ||||| ||||| |||||
Db 202 PheTyrsasnlyrlyrPheasnlyrValPhe-AspHlsMetTyrlleuPheasnhsleluI 221
QY 508 -----ATTATTCATTAACCTTATGAGAAACCAATACATTCATTCATTCACAT 551
      ||||| ||||| ||||| |||||
Db 221 eTyrlYsleuIleuPheIlepheliehuasnlystyryleuasnasnSerAsnle---Proph 240
QY 552 AGGTAGAAATTAATCTGATGATG-----GGTTATTTAAGCAATGACACACT 596
      ||||| ||||| ||||| |||||
Db 240 easnlysasnlelleeglnwecIupheasnleuTyrlTyrlPheArGlu-----11 258
QY 597 TAAACAGATTTTATTTATACATTTTGTATTAAGCAATATGATGATGATGACAGC 656
      ||||| ||||| ||||| |||||
Db 258 elYsasngluIlysasntYrllele-----LysMetAsnlyslYsgluI 273
QY 657 CTATGACTATGCTGTGAAATGTTTGCAAACTCACACCTGACCACTGACATACATTCGA 716
      ||||| ||||| ||||| |||||
Db 273 eTyrlYs-----LycCysPheAllyspne---HlsgluasnValAspHlsIleas 289
QY 717 TAATGACGTGATATA-----TATGATTTTATGATATGATGATCTTACTGAAATATGA 732
      ||||| ||||| ||||| |||||
Db 289 pasngluIlyslleleuasnilleleuArGleuTyrlValAspasnSerlleleuAspIleas 309
QY 733 -----TTAGTATGTCGCATTTTCATATATAGTATATATTCCTCAATTTTGACTA 782
      ||||| ||||| ||||| |||||
Db 309 pIleasnasnlysmetleucysasnleuasnasnleuIleasnngluasnillegluIy 329
QY 783 T---AACAAATTAACA-----TTTCATTTGATATATGACATCTTACTGAAATATGA 833
      ||||| ||||| ||||| |||||
Db 329 rIleaserlyslleleuasnphetyrlyshrlleuIlelyslYsglylYstyrlYsasnas 349
QY 834 AATGACACGTTTTCACATTCATCAACCAATATCAAGATTAATAATCTTATACACATTA 893
      ||||| ||||| ||||| |||||
Db 349 pMetThrIleTyrlYsleuIysglu-----ValIlelysalatHlsHlsH 364
QY 894 TCATTTCCATGAT-----ATGATTTTATGACATATTAATCAATC-----TATCGTGG 944
      ||||| ||||| ||||| |||||
Db 364 slleleucysAspIlyshrlYlysasnleuIuhhrPheCysSerAspIleasplySerH 384
QY 945 TGCTTTAATATGATATACACCAATACATA---AAACAACATTAATGATGAGCC---TT 997
      ||||| ||||| ||||| |||||
Db 384 rleuIleuasnSerleuasnlyshPheIleleuasnlyslleIleasplyAsnPhel 404
QY 998 GTTTTCATTTGACATCAATTCGACTA-----TC 1027
      ||||| ||||| ||||| |||||
Db 404 eleuPheTyrlglucysleuIleuIyslleleuIleuasnIlelyslPheValAspPheIlnse 424
QY 1028 CTATTCGTAT-----GTATCATGAAAAAATTCACACATGTTATACCTTTACGACACT 1081
      ||||| ||||| ||||| |||||
Db 424 rleucysIleaserleuIleuIleuIyryrlyrasnilleleuArGAsnasnva 444
QY 1082 ATTCAGAACCAACGTTAATCCCTACTCTTTTGTAGATGACATATTTTATGATATATA 1140
      ||||| ||||| ||||| |||||
Db 444 llyr-----llevalasnValleuPhe-AsnAspIleMetlyspheSerleTyrl 462
QY 1141 -----AAGATGATTAAGATATATTAACGATGATGATATTA 1177
      ||||| ||||| ||||| |||||
Db 462 eucysasnIlePheleuglylYsArgIleTyrlThclunsgluasnIleValleuIleI 482
QY 1178 TT-----AAAATTAATACGCTGATTAAC 1201
      ||||| ||||| ||||| |||||
Db 482 leHlsasnasnaspIlnhrAsnTyrlSerAsnlyslunsgluasnIleleIleIleeg 502
QY 1202 GTCAATGATTCGTAATAATAC-----TATATATAT----- 1230
      ||||| ||||| ||||| |||||
Db 502 lnlYsArgIlelysgluTyrlIlePheTyrlYsmetgluasnlyrlyslYsasPheHlsPhe 522
QY 1231 -----GATATGATTCGTTATATGAT-----ACAATATCAATTAAGATGATTC 1276
      ||||| ||||| ||||| |||||
Db 522 yslleuIysspserAspIleuSerIlelyslleuIleuSerAsnThrPheVallyslle 542
QY 1277 AAGACAT-----ACGGGTATGATTCGATG----- 1302

```

```

Db      542  singlValValyrAsnSerTyrAspPheTyrLeuLeuPheAsnAsnIleSerCysIleLeuT 562
      : : : : :
QY      1303  -----CATATACGTTGCTTTGTTATATATGAAAGTCGAATTCAT-----G 1351
      : : : : :
Db      562  yTAsnPheLeuValAsnAlgaSnsSerValLysLysTyrLysAspThrTyrIleTyrIleL 582
      : : : : :
QY      1352  CACGTGATATATATTTTCAAACTATTTTATTAAACACACAGTAAGTAAACAAA- 1410
      : : : : :
Db      582  euAsnAspSerPheValTyrLysTyrIleLysAsnAsnAspArgThrLysLysL 602
      : : : : :
QY      1411  -----ATCATA 1417
      : : : : :
Db      602  yAsnPhePheLeuLeuSerSerSerMetLysGluLeuIleCysLysAsnIleLeuSerV 622
      : : : : :
QY      1418  TSACATCACCTTACGACATACATACATATGATATGACACACACCCATCTCAATG 1477
      : : : : :
Db      622  alSerAsnArgTyrIleLysHISLeuHISGluGlaAspAsnHeasp-----GlnLys 640
      : : : : :
QY      1478  AGAGGCTATGTATCTAAAGTCGTTTAAAGCATATATGCGATACCTGCATTCGTT 1537
      : : : : :
Db      640  spGlnTyrValCysSerLeuThrPheLeuAsnAsnLeuPheAspLysIleIleHISP 660
      : : : : :
QY      1538  CCAATTTACTTATCCGTTAGATGATACAACTACATATC--ATTACG 1594
      : : : : :
Db      660  heIstYr-----IleTyrAsnLeuTyrCysHISV 670
      : : : : :
QY      1595  GTTCAAAAACACTGACATATATATTTCTCTACATTTTTCACATCAGCTTATGT 1654
      : : : : :
Db      670  alTyrLysThr-----TyrAsnTyrPheLysCysAsnLysLeuLeuAsnGluAspIleI 688
      : : : : :
QY      1655  ATATGATTTGGT-----CCTTCCATACCTTACGGAAGTGAATTGACG 1702
      : : : : :
Db      688  leSerLeuLeuLeuLeuThrCysSerLysPheGlnTyrPheIleGluAsnAsnSerAsn 708
      : : : : :
QY      1703  ACAATTTATTTATTCGATCTGATAGTTGTATATGAA-----TCCG 1747
      : : : : :
Db      708  spArg-----TyrCysArgLysGluLeuIleHISLeuLysTyrAsnIleIleAspR 726
      : : : : :
QY      1748  TTGTTAACCCCTTATGAAACCCAGTTTATTCGACCCGCTTACCTAGTAAATGGATA 1807
      : : : : :
Db      726  euLleLysAsnTyrLeuAsn-----ThrTyrLysSerIleSerIle----- 739
      : : : : :
QY      1808  TTGMAAACGACAGATAGATAGATGTTGTA--CTGATCATAGAAATATGCAATG 1864
      : : : : :
Db      740  -----AspAsnIleSerLysIlePheIleSerLeuSerAsnSerLysTyrThrCys 757
      : : : : :
QY      1865  AAGTGAATGGAAGATTAAATTCCTTCGCTGGATATACCGAAACGCTTGATACAA 1924
      : : : : :
Db      757  lValAlaSnGluAsnLeuLeuGluSer-----LeuG 768
      : : : : :
QY      1925  GCCTGATTTTGAACCTTGTACGTGAA-----CAATTCCTTTGACG 1966
      : : : : :
Db      768  lnsSerLuhGluLysValThrLysThrSerLysLysGlyLysIleHISMetLasp 788
      : : : : :
QY      1967  GTCCCTTATGAAAACATAAAGTATCTATATAGAGCAAGTACAAATATGATATATC 2026
      : : : : :
Db      788  snAsnLeuLeuAspAsnAsnSerCys-----Glu-LysTyrCgluHISArgTyrIle 805
      : : : : :
QY      2027  CGCTTAAACGTGAATTTGATGTAATGTATATGATGATATTTTACTGATGAAGCTTA 2086
      : : : : :
Db      806  GluTyrLysLysGlu-----AsnLeu 812
      : : : : :
QY      2087  ATATGAACGTCGATTTATATTAAGACGCTAGACAAAATTTGACCATATCTATTTG 2146
      : : : : :
Db      813  PheIleAsnLeuAsn-----LysIleIleGluCysLeuIleLysLeu 826
      : : : : :
QY      2147  ATATATTTCTTATATGAAAGTACGCGTTTCATTTTCATTACGATATTTCCAG 2206
      : : : : :
Db      827  AsnIlePheLeuTyrLeuLysLysLysLysThrTyrLeuTyrLeu-----TyrLysGln 844
      : : : : :
QY      2207  -----TTGACGCTCAG-----TACATAACAAT 2230
      : : : : :

```

```

Db      845  SerLeuCysProIleAsnLeuLysGluAsnIleLeuLysLysIleLeuTyrIleAlaAsn 864
QY      2231  CTGATTGCAATATAT 2245
Db      865  AsnLeuTyrMetLys 869

```

RESULT 3

```

ERBP29
DNA-directed DNA polymerase (EC 2.7.7.7) - phage phi-29
N:Alternate names: early protein gp2
C:Species: phage phi-29
A:Note: host Bacillus subtilis
C:Date: 15-Nov-1984 #sequence, revision 15-Nov-1984 #text, change 23-Jul-1999
C:Accession: A04282; B93439; S11668; S11669
R:Yoshikawa, H.; Ito, J.
Gene 17, 323-335, 1982
A:Title: Nucleotide sequence of the major early region of bacteriophage phi29.
A:Reference number: A91493; MUID:82262795; PMID:6809534
A:Accession: A04282
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-575 <10S>
A:Cross-references: GB:V01155; GB:J02478; NID:915659; PUDN:CAA24480.1; PID:915667
A:Note: the authors report the amino acid composition of the translated protein; the
R:Escarim, C.; Salas, M.
Nucleic Acids Res. 10, 5785-5798, 1982
A:Title: Nucleotide sequence of the early genes 3 and 4 of bacteriophage psi29.
A:Reference number: A93439; MUID:83064518; PMID:6292852
A:Accession: B93439
A:Molecule type: DNA
A:Residues: 1-85 <ESC>
C:Genetics:
A:Gene: 2
A:Map position: 16-6
C:Superfamily: phage PZA DNA-directed DNA polymerase
C:Keywords: DNA binding; DNA replication; early protein; nucleotidyltransferase

```

Alignment scores:

```

Pred. No.: 2,22e-07 Length: 575
Score: 241.50 Matches: 151
Percent Similarity: 35.57% Conservative: 114
Best Local Similarity: 20.27% Mismatches: 239
Query Match: 5.93% Indels: 241
DB: 1 Gaps: 40

```

US-09-727-892A-2 (1-2286) x ERBP29 (1-575)

```

QY      31  AAGATGACGTCGAATGATTTTATACGGGATATAGAAACATATAGCGTACAATAAGTT 90
      : : : : :
Db      2  LysHISMetProArgLysMetTyrSerCysAspPheLuhR----- 15
      : : : : :
QY      91  AACGACGMAAAACCAACCAATATATAAAACCTTACTATTCGTAGCAATTTGGTTG 150
      : : : : :
Db      16  -----ThrThrLysValGluAspCysArgVal-----Tyr 25
      : : : : :
QY      151  TTTATGCTTATGAAATTTGATGTTGATTTCCGAGTTTCGAA-----TCCTTTTAT 204
      : : : : :
Db      26  AlaTyrGlyTyr---MetAsnIleGluAspHISerLeuTyrLysIleGlyAsnSerLeu 44
      : : : : :
QY      205  GACGCAATTTTATAGTATGTGMAAAGACGTGATCAATCAACAAATCAAAACAGATATT 264
      : : : : :
Db      45  AspGluPheMetAlaIrrVal-----LeuLysValGlnAlaAspLeu 58
      : : : : :
QY      265  ATCATGATTCGACATACATGTAATTAATATAGATATCATTTTATTTACTTAAAGACACATG 324
      : : : : :
Db      59  TyrPhe-----HisAsnLeu---LysPheAspLysAlaPheIleIle----- 71
      : : : : :
QY      325  CGTATATTTGATATATATATACACGCGGAAAATATATATTAATATCGACGAAGAAATGAA 384
      : : : : :
Db      72  -----AsnTyrPheGluArgAsnGlyLysPheLysTyrPheSerAlaAspGly----- 85
      : : : : :
QY      385  CACACATTMAAAATGAAAGAGGCTACTATTTTACCCAAAATCAAATGTATTTTGAAG 444
      : : : : :

```


Db 2414 LysLysLysMetPheHisLeuGlnIleSerGluTyrAspLysHisThrAsnTyrAsnSer 2433
|||
QY 790 TTAACATTTCATGGAATATATGAGATCTTACTGGAA--AATGAATGCACAGCTTT 846
||| |||

Db	2434	Leu-----	-pheketaspCysValGlnAsnHisAsnIleIysIysMetAsnSerThr	2451
Qy	847	CAGTTACTCAACCAACATATCAGAGATTATTAATAATCTTATACACATTATCAT	-----	897
Db	2452	AsnAsnMetAsnHisHisIleAsnThrAsnAsnAsnTyrLeuHisAsnHisAsnPhel	2471	
Qy	898	-----TTCCATGATATGATTTTTTTATAGCTATATTAATTCATTCTTATACGGTGGTGTTA	951	
Db	2472	SerAsnTyrAsnSerPheAsnValHisAspAsn--LysLysIleTyrSerTyrAsnGlu	2490	
Qy	952	AATATGATATACACCAATATCATTAACCAACATATATGAGACCCCTGTTTCTATTCAC	1011	
Db	2491	AsnGlySerSerAspGluIleMetGlnIysIysIleAspMetSerIleTyrIys--Asn	2509	
Qy	1012	ATCATATGAGAGTATTCCTTATGTGATGTAT--CATGAAAAATTCACACATGTTATAC	1068	
Db	2510	IleAspSerIlePheProGluThrPheIleAspSerAspLysGlnProAla-----Tyr	2527	
Qy	1069	TTTATACGACACATATCAGAACCAACAGGTATATCCATCTTTTATGATGAGACATATAT	1128	
Db	2528	AsnPheAspProIleAspSerIleAsnLeuGlySerSerArgSerAsnGlnIuIysIys	2547	
Qy	1129	TTTTCATTATATATAGATTGTGATTAAGATATTTTACGATGATTTATTAATTAATTAATAA	1188	
Db	2548	LysLysTyrIleGlnIleAspAsnProValIysGlnCysLeuLeuLeuAsnIleAsn	2567	
Qy	1189	TCACGTGATTATACCTCAAAAGATTGTAATAATACAT	-----	1224
Db	2568	TyrAspLysHisAspSerIleValTyrAsnLysTyrAspAsnMetPheHisTyrAspGlu	2587	
Qy	1225	-----AATATGATATATGATATACGTATATATCATATCAATACATTAATAGA	1269	
Db	2588	LeuProAspIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAspAsnAsnAsnThrCys	2607	
Qy	1270	ATGATTCAAGACATTTACGGGATTTGATGCATGCATATACGCGTAAATTCGTTGGTTATA	1329	
Db	2608	ValIleGlnAspIleIys-----AspIleTyrGlnLysArgMetAsnLysAsnThrLys	2625	
Qy	1330	TATGATAGTGAATCTTCATGCACGAGATATATTTCAACATATTTATTTATTAACA	1389	
Db	2626	ArgAsnLysGlnIuIysLysGlnLysArgLysTyrIlePheLeuAsnAsnPheAsnAsnAsn	2645	
Qy	1390	CAGGTAAGTTAAAAACAATAATATATGATCATCACTTACGACTATCAC-----	1440	
Db	2646	LysGlnLys--LysMetLysAsnAsnGlnLysThrValTyrSerAsnAsnAsnIleMet	2664	
Qy	1441	-----ATTACTGATGATATCAACGAAACCCGACTCATCAATGAGAGAGGTATGTTA	1491	
Db	2665	GlyGlnGluPheTyrAsnGluPheTyrLeuHisAsnPheLysAsnGlnIleIysCysMet	2684	
Qy	1492	TCTAAAGTCGTTTA--AATGATATATATGCGATACGT-----GCATTACGTTACAT	1542	
Db	2685	LysTyrIleAsnLeuThrGlnSerLeuTyrAspValIysTyrArgLeuLeuLeuPhe	2704	
Qy	1543	TTTAACTTATCCGTTAGATGATATACAAATGAACATATCAATATCATTAACGGTTACAA	1602	
Db	2705	TyrLysPheIleIleIleLeuLysHisLysGlnLeuGlnAsnGluAsnTyrIleLys	2724	
Qy	1603	-----AACACGTGAACGTAAATATTTATCTCTACATTT	1635	
Db	2725	GlnGluLysGluPheLeuLysLysHisHisIleLysLysAsnIleProPheLeuPhe	2744	
Qy	1636	GTCACATACGCTCATCTGTATACTTATATGCGTTCCCTTCGAACTACTTAACGAAAGTGA	1695	
Db	2745	Ile-----TyrGluLeuMetIleThrPhe--PheAsnThrAlaGluAsn	2758	
Qy	1696	ATTGACGACAAATTTATTTATTTGCGATGATGATGTTGTTATATGAATACCGTTGTTAA	1755	
Db	2759	IleAsnLysAsn-----ThrTyrTyrTyrValLeuIleIleAsn	2771	
Qy	1756	CCCTTATTAACCCCACTTATTCGACCCGATAGCCTTAGTAATGGCATATTT-----	1809	
Db	2772	IleLeuValAsnLeuPheLeuPhe-----IleAsnLysArgAsnTyrAspAsp	2787	

QY 1810 -----GAAACGACAGATAGATAGATGTTGTA 1839
 Db 2788 GluThrCysMetSerAsnIleIleAsnAsnAspAsnAsnLysLysAsnLysAsnAsnLeu 2807
 QY 1840 CTGAATCTAAGAATATGCAATATGCACTGATGAAGAAAGATTAATAATGCTCTGCTGCT 1899
 Db 2808 IleGluAsnLysAsnGluIleLeuAsnThrAsn-----IleLysSerLeuLys 2823
 QY 1900 ATACGAAACGCGCTTATGACAGCGCTTATGAACTTTGAACTTTGACGGAACATTC 1959
 Db 2824 AsnAspLysGluLysIleAspAsnHisSerAsnLysIleMetPheLysCysAspLeuPhe 2843
 QY 1960 TTGACGCTGCCATTTATGAAACATTAAGATCTATAATGAGCAAGGATGACATA--- 2016
 Db 2844 CysAspAspPhePheIleSerAsnGlyLysLys-----AsnLysGluAsnValIlePhe 2861
 QY 2017 ---TCGATATATP-----CCGCTAAACTGAAATTTGATGCTGTAAT 2055
 Db 2862 HistHleuHisAsnMetSerHisLysGluMetSerLysLysAspLeuIleGlyLysAsn 2881
 QY 2056 GTATATGATGAATATTTTACTGATGAACTTAATATGAACGTAATTAATTAAGAC 2115
 Db 2882 LysLysLysGluAsnLysIleAsnAsnLeuIleLeuGluLysLysLysLysLysAsn 2901
 QY 2116 GCTAGAGAAATTTGACCATATGCAATTTGATATTTCTTTATTTGAAAGTGACATC 2175
 Db 2902 LeuAsnValHisIleAsnLysLysMetLysAsnAsnIleLeuLys----- 2916
 QY 2176 GGTTCATTTTCATCAGACTATTTCCAGTTGACGCTCAGATACATCAATCTGAT 2235
 Db 2917 ---SerPhe---IleAsnArgIleAsnGluThrArgAspAsnThrLysLysAsnLys 2934
 QY 2236 TTGCATATA-----TTAAACGTGACATGATGAAATTAAGAAAGCGAAC 2280
 Db 2935 LeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2950
 RESULT 5
 J00161
 DNA-directed DNA polymerase (EC 2.7.7.7) - phage M2
 C:Species: phage M2
 C:Date: 07-Sep-1990 #sequence-revision 07-Sep-1990 #text-change 20-Sep-1999
 R:Accession: J00161
 R:Matsumoto, K.; Takano, H.; Kim, C.I.; Hirokawa, H.
 Gene 84, 247-255, 1989
 A:Title: Primary structure of bacteriophage M2 DNA polymerase: conserved segments within
 A:Reference number: J00161; M01D:90128268; F01D:251515
 A:Accession: J00161
 A:Molecule type: DNA
 A:Residues: 1-572 <MAT>
 A:Cross-references: GB:K33144; NID:q215507; PIDN:AAA3366.1; PID:q215509
 C:Comment: This enzyme catalyzes the formation of a primer protein, 5'dAMP initiation co
 C:Gene: G
 C:Superfamily: phage P2A DNA-directed DNA polymerase
 C:Keywords: DNA binding; nucleotidyltransferase
 Alignment Scores:
 Pred. No.: 5,89e-07 Length: 572
 Score: 234.50 Matches: 136
 Percent Similarity: 35.33% Conservative: 97
 Best Local Similarity: 20.81% Mismatches: 221
 Query Match: 5.76% Indels: 211
 DB: 2 Gaps: 32
 US-09-727-892a-2 (1-2286) x J00161 (1-572)
 QY 382 GAAACACATTAATAAAGAGAGCGTACTATTTAGCCCAAAATCAAAATGTAATT--- 438
 Db 11 GluThrThrThrLysLeuAspAspCysArgValTTPAlaLysGlyLysMetGluIleGly 30
 QY 439 -----TTAGAAAACGTTGTAATCTCAATCAAT 468

Db 31 AsnLeuAspAsnLysLysIleGlyAsnSerLeuAspLysLeuPheMetGluIntPValMetGlu 50
 QY 469 TTACATTTAACATGATGTTTAAATGTTTAAATTAAT-----ATTATGATTAACCTT 522
 Db 51 IleGluIleAspLeuLysPheHisAsnLeuLysPheAspGlyAlaPheIleValAsnTrp 70
 QY 523 ATGAAAC-----AATACATCA 540
 Db 71 LeuGluIleHisGlyPheLysTrpSerAsnGluGlyLeuProAsnThrLysAsnThrIle 90
 QY 541 ATTGCAACATTAAGTAGA---AATTACTGAT-----GGTGATATTAAACGATCA 591
 Db 91 IleSerLysMetGlyGlnTrpLysMetIleAspIleCysPheGlyLysGlyLysArg 110
 QY 592 CAACCTTAACA----- 603
 Db 111 LysLeuHisThrValIleLysAspSerLeuLysLysLeuProPheProValLysLysIle 130
 QY 604 -----GATTTTATATATACGATTTTATGATTAAGATATGATATGATAGTGAAGCC 657
 Db 131 AlaLysAspPheGlnLeuProLeuLeuLysGlyAspIleAspLysTrpHisThrGluArgPro 150
 QY 658 TATGACATCTCTGGAATGTTTTCGAAAACCTACACCTGGAACACTTACATACATTCAT 717
 Db 151 ValGlyHis-----GluIleThrProGluLysLysLysLys 164
 QY 718 AATGACGTATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
 Db 165 AsnAspIleLeuIleIleIleAlaArgAla-----LeuAspIleGlnPheLysGln 180
 QY 778 GACTATATACAAATTAACATTTTCATGATATATGATGATGATGATGATGATGATGATGAT 837
 Db 181 GlyLeuAspArgMetThrAlaGlySerAspSerLeuLysGlyPheLysAspIleLeuSer 200
 QY 838 ACACGT-----TTTCAGTTACACCAACATATCAACATATTAATAATCT 882
 Db 201 ThrLysLysPheAsnLysValIleProLysLeuSerLeuProMetLysLysGlu----- 218
 QY 883 TATACACATTTATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 942
 Db 219 -----IleArgLysAlaLysArg 224
 QY 943 GGTGTTTAAATATGATATACACCAATACATTAACAACTAATGATGACCTTCTTTT 1002
 Db 225 GlyGlyPheThrTrpLeuAsnAspLysLysGlyLysGluIleGlyLysLysLysLysLysLys 243
 QY 1003 TCTATTGACATCAATTCGAGTTATCCTTATGATGATGATGATGATGATGATGATGATGAT 1062
 Db 244 ValPheAspValAsnSerLeuLysProSerGlnMetLysSerArgProLeuPro----- 261
 QY 1063 TTATACTTTTACACACATTTTACAGACCAACGCTTAATCCTTACTTTTAAATGATGAC 1122
 Db 262 -----TyrGlyAlaProIleValPheGlnGlyLysLysLysAsp 275
 QY 1123 AATTATTTTCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1173
 Db 276 GluGlnLysProLeuLys-----IleGlnArgIleArgPheGluPheGluLeuLysGlyGly 294
 QY 1174 -----TTAATTAAATTAATCAGCTGATTAAGTCAATGATGATGATGATGATGAT 1224
 Db 295 TyrIleProThrIleGlnIleLysLysAsnPro-----Phe 306
 QY 1225 AATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1284
 Db 307 PheLysGlyAsnGluLysLysLys-----AsnSerGlyValGluProValGluLeuLysLys 325
 QY 1285 ACGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1344
 Db 326 ThrAsnValAsp-----LeuGluLeuIleGlnGlnHisLysGluLeuLysAsnValGluLys 344
 QY 1345 -----TTTCATGACAGCTGATATATTTTTCAAACATAT----- 1377
 Db 345 IleAspGlyPheLysPheArgGluLysTrpGlyLeuPheLysAspPheIleAspLysTrp 364

```

QY 1378 ---TTTATTAACACAGAGTAAAGTTAAAAACAAATCATATGACATCAGCTTACGAC 1434
      :|||||:|||||:
Db 365 ThrTyValLysThrHisGluGluValLysLys----- 376
QY 1435 TATCATATGATGATGATATCAACAGACACCATCTACTCAATAGAGAGTTATCTTACT 1494
      :|||||:|||||:
Db 377 -----GlnLeuLys 379
QY 1495 AAGTCGTTTAAATGATTAATATGAC----- 1521
      :|||||:|||||:
Db 380 LysLeuMetLeuAsnSerLeuTyrGlyLysPheAlaSerAsnProAspValThrGlyLys 399
      :|||||:|||||:
QY 1522 ATACCGCATTTGCTGATCATTTTAACTTA---TTCCGTTTATGATGATACATGACATGA 1578
      :|||||:|||||:
Db 400 ValProTyrLeuLysAspAspGlySerLeuGlyPheValGlyAspGluGlu----- 417
      :|||||:|||||:
QY 1579 TACATATGATTAACGCTTACAAAAC-----ACTGAACGTAATATATTA----- 1623
      :|||||:|||||:
Db 418 -----TyrLysAspProValTyrThrProMetGlyValAlaPheIleThr 431
      :|||||:|||||:
QY 1624 -----TTCTCTCATTTGTGCATCATCGCTTCATTTGATTAATCTATGCTTCT 1671
      :|||||:|||||:
Db 432 AlaTyrAlaArgPheThrThrIleThrAlaAlaGlnAlaCysTyr----- 446
      :|||||:|||||:
QY 1672 TTCCATATCTTACGGAAGTAAATTGACGACAAATTTATTTATTTGCGATGATAGT 1731
      :|||||:|||||:
Db 447 -----AspArgIleIleTyrCysAspThrAspSer----- 456
      :|||||:|||||:
QY 1732 TTGTATATGAAATCGTGTGTTAAACCTTATGAAACCCAGTTTATGACCCGATAGCC 1791
      :|||||:|||||:
Db 457 IleHisLeuThrGlyThrGluValProGluIleIleLysAspIleValAspProCysLys 476
      :|||||:|||||:
QY 1792 TTAGGTAATGGGATTTGTAAGAACAGACAGATGATGATGTTGATGATCATAG 1851
      :|||||:|||||:
Db 477 LeuGlyTyrThrAlaHisGluSer---ThrPheLysArgAlaLysTyrLeuArgGlnLys 495
      :|||||:|||||:
QY 1852 AAA-----TATGCATATGAGTGAATGAGAAAGATTAATTAATCTCTCTGCT 1896
      :|||||:|||||:
Db 496 ThrTyrIleGlnAspIleTyrValLysGluValAspGlyLysLeuLysGluCysSer--- 514
      :|||||:|||||:
QY 1897 GGTATACCGAAAGCGCTTGTATACAGCGTCGATTT-----GAA 1938
      :|||||:|||||:
Db 515 -----ProAspGluAlaThrThrTyrLysPheSerValLysCysAlaGlyMetThrAsp 532
      :|||||:|||||:
QY 1939 ACCTTTGTACGTGAACAAATCTTTGACGGTGCCTATTTGAAAACAAATAAAGTATCTAT 1998
      :|||||:|||||:
Db 533 ThrIleLysLysValThrPheAspAsnPheAlaVal----- 545
      :|||||:|||||:
QY 1999 AATGACACAGGTACATATGATATATCGCTTAAGACGAATGTATGTTATGTA 2058
      :|||||:|||||:
Db 546 -----GlyPheSerSerMetGlyLysProLysProValGlnValAsnGlyGlyVal 562
      :|||||:|||||:
QY 2059 -----TATGATGATATTTTACT 2076
      :|||||:|||||:
Db 563 ValLeuValAspSerValPheThr 570
      :|||||:|||||:

```

```

A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0375w

Alignment Scores:
Pred. No.: 1,09e-06 Length: 1802
Score: 230.00 Matches: 202
Percent Similarity: 33.77% Conservative: 131
Best Local Similarity: 20.49% Mismatches: 242
Query Match: 5.65% Indels: 411
Db: 2 Gaps: 59

US-09-727-892a-2 (1-2286) x G71616 (1-1802)

QY 7 TTACAGATGATGATGATATCATATAA-----CATGAA 39
      :|||||:|||||:
Db 931 LeuTyrLysCysLeuGlnTyrLeuSerLysLysAsnAspLeuThrLeuPheIleAsnGlu 950
      :|||||:|||||:
QY 40 CGTGAATGATTTATATCTGGATATA---GAAACATTTAGCGTACATTAAGTAAACGA 96
      :|||||:|||||:
Db 951 ThrGluIleIleMetTyrLeuAsnIleValLysLysLeuLysGluArgLysIleAsnAsn 970
      :|||||:|||||:
QY 97 CGAAAAAACCAACCAATATATAAAGTT---ACTATATCTGACCAATGTTGTT 153
      :|||||:|||||:
Db 971 IleAsnGlu-----LysPheLysAsnIleProAsnHisLeuLysGlnIleLysGluIle 988
      :|||||:|||||:
QY 154 AATGATTATGAATGATGATGATGATATTCGAGTTTCGATCTTTTATGACGCATTT 213
      :|||||:|||||:
Db 989 LysGlnValLysGluAspIleLeuGlnAspLysAsnThrLysAsnIleTyrGlnMetIle 1008
      :|||||:|||||:
QY 214 TAT-----ACGTATGTGAAAGAGCTGATACAAACAAATCA--- 252
      :|||||:|||||:
Db 1009 HisAsnTyrGlnThrAsnIleThrTyrGlnThrLysAsnGlnAlaValThrProSerCys 1028
      :|||||:|||||:
QY 253 -----AAACAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 297
      :|||||:|||||:
Db 1029 CysTyrHisAsnThrSerHisIleIleLeuAsnThrHisGlu---AsnIleTyrGluGlu 1047
      :|||||:|||||:
QY 298 -----AATCATTTTCTTAAAGACACCATG-----CGTTAT 330
      :|||||:|||||:
Db 1048 LysLysLysAsnAsnValLeuLeuAsnAspAspLeuTyrAspGluIleLeuGluArgTyr 1067
      :|||||:|||||:
QY 331 TTGATATATTT----- 342
      :|||||:|||||:
Db 1068 MetAsnLysIleMetLeuAspAsnLeuPhePheSerSerPheGlnLysValGlyLysLys 1087
      :|||||:|||||:
QY 343 ---ACACCGCAATATA-----TATTTAAATCTCGAGAGAAAT 381
      :|||||:|||||:
Db 1088 TyrThrHisTyrPheAsnLeuSerSerSerLeuIleGlnTyrAsnLysIleLeuGluGluAsn 1107
      :|||||:|||||:
QY 382 GAACACACATTAATAATGAAGAGCTACTATTTTACGCCAAATAATCAAAATGTATTTTA 441
      :|||||:|||||:
Db 1108 Lys-----LysAspLysThrIle-----AsnAsnGluAsnAspIleIle 1120
      :|||||:|||||:
QY 442 GAA-----AAACGTGTTAATCTTCAATCATTTAGAT----- 474
      :|||||:|||||:
Db 1121 LysIleAspAsnAsnLysAsnGluGlnSerIleAsnValAspAsnMetTyrThrSerSer 1140
      :|||||:|||||:
QY 475 ---TTAACAATGTTT-----TTAATGTTTAA----- 501
      :|||||:|||||:
Db 1141 LysCysThrLysPheProPheAsnIleHisAspPheLysLysTyrSerIleAsnIleTyr 1160
      :|||||:|||||:
QY 501 ----- 501
      :|||||:|||||:
Db 1161 PheLeuValTyrAspAsnIleLeuSerTyrAsnLysLysIleAsnLysGluGluIleGlu 1180
      :|||||:|||||:
QY 502 -----TTTATATATATGATATATCTTTTGAAGAACCATATCATATGACAAACA----- 549
      :|||||:|||||:
Db 1181 LysIleTyrPheAsnIleLeuAspAsnMetIleLysTyrLysGlnAsnValLeuThrGluAsp 1200
      :|||||:|||||:
QY 549 ----- 549
      :|||||:|||||:
Db 1201 AsnPheTyrTyrIleIleSerAlaLeuLeuLysAlaGlnAsnPheGluHisGluValTyr 1220
      :|||||:|||||:

```


C:Accession: G71607
 R:Gardner, M.J.; Jettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: G71607
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1182 <GAR>
 A:Cross-References: GB:AE001414; GB:AE001362; NID:g3845260; PIDN:AAC71936.1; PID:g384526
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0735c

Alignment Scores:
 Pred. No.: 1,18e-06 Length: 1182
 Score: 229.50 Matches: 185
 Percent Similarity: 35.96% Conservative: 135
 Best Local Similarity: 20.79% Mismatches: 287
 Query Match: 5.64% Indels: 283
 DB: 2 Gaps: 46

US-09-727-892a-2 (1-2286) x G71607 (1-1182)

```

QY      25 TATCATTAACATGAACGTGCAATGATTTTACTGGGATATAGAAACATTAACGTCAT 84
      ::::|||||:  |||  ::|||:|||||:  |||  |||
Db     141 Phasnllysasnary-----IleTyPheaspLeuHsValLeuPheLysasn 157

QY      85 AAAGTATACGACGAGAAAAACCAACCAATATATAAAGCTTACTATCTGTAGCAAT 144
      ::::  ::::  :|||:|||||:  |||  |||
Db     158 AspleuleuleuInlaryasn-----IleasnIleSerTyPheLysasn 173

QY      145 GGTTGGTTTAAAT-----GGTTATGAATTTGATGTAAGTATTTCCGAGTTTGCATCT 198
      ::::  :|||  :|||  :|||:  :|||:  :|||
Db     174 AspasnMetSerArgLuglValHsHsLysArgAspIleuleuLeasnHsGlnCys 193

QY      199 TTTTATGAC-----GCATTTTATACGATTTGGAAGA----- 231
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     194 LeuTyrsnIleasnspLeuPhealaLeuPheIlePheTyrsValHsIleLysArgPhe 213

QY      232 -----CGTGCATCAATCAAAAATCAAAAACAGATTTTTCATGATTCACAT 279
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     214 TyPheaspPhePhePheHsIleLeuLysasnIleasnAspMet-----Glu 229

QY      280 AACTGTATATAATCGATATCATTTTCTTAAAGACACCATGGGTTATTTTGCATAT 339
      ::::  |||  |||  |||  :|||  :|||  :|||
Db     230 SerThrnsnAspTyLysasn-----ValCysTyPheLysasn 242

QY      340 ATTACACGCGAATAATATTTTAAATCGACAGAAAT----- 381
      |||  :|||:  :|||:  :|||:  :|||:  :|||
Db     243 IleHsLysGlnHsIleTyHsIlePheProHsLysasnTyTyrsnIleGlnasn 262

QY      382 -----GACACACATTTAAATGAAAGAGGCTACTATTTTACCAAAAT----- 426
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     263 MetasnSerLutyrCysLeuLysPheLeuLysAlaCysIleGlnLeuLysasnIleIle 282

QY      427 CAAATGTATAT-----TTAGAAAACGTGTAACTTAACCTTAATCAAT 468
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     283 SerasnIleValasnIleasnLysLysLysGlnLysasnValHsHsIleGlnasn 302

QY      469 TTGATTTAACAATGTTTAAATGTTTAAATTTAATTTATTTAGTAACTTTATGAA 528
      |||  :|||:  :|||:  :|||:  :|||:  :|||
Db     303 AsnIleArgThrCysArgIleasnTyPheValPhe-----IleLysasnAlaIlePhe 320

QY      529 ACCAATCATCAATTCGACACATAGTAAATTTACTTGATGGTGTATTTAACAGAA 588
      |||  :|||  :|||  :|||  :|||  :|||
Db     321 LysLysCysLysLys-----IleLysLysLys----- 329

QY      589 TCACAACTTAACACAGATTTTAAATATACATTTTGAATTAAGATATGATATGAT 648
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     330 -----GluLysLysLysLysLysasnAsp 337
  
```

```

QY      649 AGTGAACCTATGACTATGCTGCAAAATGTTTGCAAAACACACCTGGAACAATTACA 708
      ::::  |||  |||  |||
Db     338 GluGlnIleTyLysLysAla----- 344

QY      709 TACATTTCAATGACGTCATTTATTTAGCTATGTCACATATTCATTTTACGATATTT 768
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     345 TyrlleHsasnSerVal-----TyrlleHsIlePhe 355

QY      769 CCA-----AATTTGACTATACAAATTAACATTTTCATGCAATATATG 813
      :|||  :|||  :|||  :|||
Db     356 LysasnMetLeuLeuHsIleasnIleLysIleGlnArgLysLysLysIleasnAsn 375

QY      814 GAATCTTACTGAAATATGAAATGACACGTTTTCAGTTACTCAACAAATCAAGAT-- 870
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     376 AsnLysIleIleasnLysIleIleLysasnIleIleGlnLeuPheasnAsn 395

QY      871 ---ATTAAATATCTTATACACATTTATTC-----CATGATATG 909
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     396 IleIleArgLysLysTyrlleHsPhePheLeuLysGlnLysTyrlleLysMet 415

QY      910 AATTTTATGACTATATTAATCAATTCATCGTGGTGTATTAATATGATTAACACCAA 969
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     416 ThrTyrsLysPheLysLysArgLysAspMetasnThrLeuIleMetCysAsp---Lys 434

QY      970 TACATTAACAACATATGATGAGCCTTGT---TTTCTATTCACATCAATTCGATAT 1026
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     435 TyrlleasnLysSerIle-----CysLeuPheLeuasnAsnPheGlnAspSerSer 451

QY      1027 CCTATGATGATGATATGAAATAATTCACATGTTATCTTTTACGAACTATATCA 1086
      :|||:  :|||  :|||  :|||
Db     452 IlePheIleLysTyPheMetLysIleIle----- 460

QY      1087 GAACCAACGTTAATCCCTACTTTTATGATATGACAAATTTT-----TCATTTAT 1140
      :|||  :|||  :|||  :|||:  :|||:  :|||
Db     461 LysLysAlaasnIleIleasnTyLeuTyrsPheLysValPheIleLysSerLeuMet 480

QY      1141 AAGATTGATTAAGAT-----GTATTAAACGATTTTATTAATTAATTAATAA 1188
      :|||  :|||  :|||  :|||:  :|||:  :|||
Db     481 LysCysValLysLysAsnCysAlaTyPheThrArgLysAsnLeuPheIleTyLys 500

QY      1189 TCACGTGTA-----TTACGTCAAAATGTTGTAATTTACTATAT 1227
      :|||  :|||  :|||  :|||:  :|||:  :|||
Db     501 TrpLysThrHsMetasnLysasnLeuaspasnIleasnGlnHsasnAsnLysTyLysasn 520

QY      1228 AATGATTAATGATTAAGTATATATCAATCAATACATTAAGA----- 1269
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     521 LysHsIleasnAsnMetTyrlleLysThrAspLysValLysAspAsnValLeuPhe 540

QY      1270 -----ATGATTCAGACATTAACGGGTATGATTCATG---CATATCGGTATAT 1317
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     541 ProPheSerLeuIleLysAsp-----AspIlePheArgHsIleGlu----- 554

QY      1318 TCGTTGTTATATATGATATGATCAATCTTCATGACGCGTATATTTCAAACTAT 1377
      :|||  :|||  :|||  :|||:  :|||:  :|||
Db     555 -----AspTyrlleHsPheHsIleHsIleLysAspIleIleTyrlleCysTyPhe 569

QY      1378 TTTATTAACACAGATTAAGTAAATAAACAATAATGATACATCACTTACGACTAT 1437
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     570 -----LysasnLysLeu-----TyrlleGluTyPhe 576

QY      1438 CACATTTACTGATGATATCACACACCAATCACTCAATTAAGAGAGGATATGTTCTTAA 1497
      :|||  :|||  :|||  :|||:  :|||:  :|||
Db     577 LysLeuPheHsLysIleIleasnHsIleuIleasnIleasnLysIleCysSerLys 596

QY      1497 ----- 1497

Db     597 TyrlleValThrIleIleIleLeuLeuTyrsnLysLeuasnCysLysThrGlnLeuLys 616

QY      1498 -----GTGGTTTAAATGATTAATGATGACATACCTGATTCGTTGCATAT-- 1542
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     617 GluLeuLeuPheIleLeuLeuasnAsnTyArg-----ProSerLeuLysGlnArgasn 634

QY      1543 -----TTTACTATTCGTTTATGATGATTAACATGAACTATATC--- 1581
  
```

Thu Jan 9 14:50:15 2003

us-09-727-892a-2.rpr

Page 11

D
635 LysArgAsnAsnIleSerIleAsnAsnIleTyrGluLeuLysAsnIleAsnLysLysTyrlle 654

OY 1581 ----- 1581

D
655 LysLysLysLysLysLysLysTyrlleTyrIleTyrIleThrIleCysLysLysLysAsn 674

OY 1582 AATATCATTAACCGTTACAA-----AACACTGAACGTATATATA 1623

D
675 AsnValGlyAlaIleHisLysHisAsnValMetMetThrSerAsnHisAsnAsnIleleu 694

OY 1624 TTCCTACATTTCGCACATCAAGTCATTGTATACCTATATGGTCCCTTCGAACTACTA 1683

D
695 PheArgSerPheGluTyrVallys---ValHisLysLeuLeuPheIleAsnIleleu 713

OY 1684 ACGGAAGTGAAATGACAGCACATTTATTATTGCGACTGATAGATTGTATATGAAA 1743

D
714 IleLysSerAsnIleTyrIleAsnTyrGluTyr-----SerLeuTyrPheleu 729

OY 1744 TCCGTTGTTAA-----CCCTTATGAACCCAGCTTATATGACCOCGATACCTTA 1794

D
730 SerLeuIleLysGlnLysHisAlaPheIleLysLysLysGlyPheTyrIleleucysTyr 749

OY 1795 GGTAAATGGATATTGAACACGACAGATA-----GATAGATGTTTGT 1839

D
750 IleLeuPheHisIleGlnAsnAsnHisIleIleTyrLysSerTyrGluHisIlePheAsn 769

OY 1840 CTGATCTATAGAATATGCAATGCAATGCAATGSAAGATTTAA----- 1884

D
770 ProTyrAsnLysTyrAsnIleTyrAsnIleTyrAsnIleIleLysCysThrLeuProGln 789

OY 1885 -----ATTGCTTCGCTGTATACCGAAA 1908

D
790 IleleuGlyThrSerAsnIleTyrSerleuIleTyrValAlaIlePheLeuTyrSerThrAsn 809

OY 1909 AACGCCTTTGATACAGCGCTGCATTTGAACCTTTGACSTGACAACAATCTTTGACCGT 1968

D
810 AsnThrIleAsnPheIleLysIlePhePheThrIlelle---GlnLysPheTyrAsper 828

OY 1969 GCCATT-----ATGAAACACATAAAGTACTATATGACGACAGGACATATACG 2019

D
829 SerMetIleLysGlnIleGlnAsnSpLysAsnAsnTyrGlnHisIleSerCysHisAsn 848

OY 2020 ATATATCCGCTTAAACTGAATGTATGTATGATGATGATGATATTTACTGAT 2079

D
849 TyrSerProLysAspAsn-----SerGluTyrTrilepro 861

OY 2080 GAACCTATATGAACAGCGATTTTATATAAAGCGCTAAGAAAAATTTGGACATAGT 2139

D
862 AspAspHisAsnLysLeuLeuTyr-----AsnTyrSerTyrAsn 874

OY 2140 CAATTGTGATATATCTTTATATGAAGTAGACATCGCTTCATTTTCTACTAACGACTTA 2199

D
875 GlnLeuTyrGluLysAsnHisPheAsnAspAsp-----AsnIlePheIleHisAspLeu 892

OY 2200 TTTCACGTTGAACGTTCCAGTACATACAAA 2229

D
893 LysIleTyrGluArgAsnIleAsnAsnLys 902

RESULT 8

ERRBPZ

DNA-directed DNA polymerase (EC 2.7.7.7) - phage PZA

A:Alternate names: gene PZA

C:Species: Phage PZA

A>Note: host Bacillus subtilis

C>Date: 30-Sep-1987 #sequence.revision 30-Sep-1987 #text.change 23-Jul-1999

C/Accession: D24528

R:Places: V.: Vicek, C.; Urbanek, P.; Hostomsky, Z.

Gene 38, 45-56, 1985

A>Title: Nucleotide sequence of the major early region of Bacillus subtilis phage PZA, a

A/Reference number: A91538; MUID:86056991; PMID:3934048

A/Accession: D24528

A/Molecule type: DNA

A:Residues: 1-572 <PAC>
A:Cross-references: GB:M1813; GB:M13904; GB:M13905; NID:q216046; PIDN:AA88478.1; P
C:Genetics:
A:Gene: 2
C:Superfamily: phage P2A DNA-directed DNA polymerase
C:Keywords: DNA binding; early protein; nucleotidyltransferase

Alignment Scores:

Pred. No.:	1,366-06	Length:	572
Score:	228.50	Matches:	127
Percent Similarity:	36.76%	Conservative:	98
Best Local Similarity:	75%	Mismatches:	184
Query Match:	5.61%	Indels:	203
DB:	1	Gaps:	34

US-09-727-892A-2 (1-2286) x ERBP22 (1-572)

```
QY 490 AATGGTTTAAATTAATTAATTAATTTGATGAACCTTATGAAGAAC---AATACATCAATTGCA 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 74 AaNgLyPhEluYrIpSer---AlaSpGlyLeuProAsnThrTyraSnhrlleIleSer 92
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 547 ACATTAGGTAG---AAATTACTGAT-----GGTGGTTATTTAACAGAAATCACAACTT 597
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 93 ArgMeGlyGlnTrpIlyrMetIleAspIleCysLeuGlyTylGlySgIlySthGlySile 112
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 AAAACA----- 603
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 113 HIsThVaIlleTyraSpSerLeuLysLeuProPheProValLysLysIleAlaLys 132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 604 GATTTTAAATTAATTAATTAATTTTGTATGAAGATATGATGATGATGATGAGCCATGAC 663
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 AsPhelYsLeuThrValLeuLysGlyAspIleAspTyrHisLysLysIlyrProValGly 152
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 664 TATGCTGTGAATGTTTTCGCAAAACTCACACCTGAAACAATTACATACATCATATGAC 723
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 153 Tyr-----GluIleThrProAspGluTyrAlaTyrIleLysAsnAsp 166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 724 GIG---ATTATATTAGGATGTGCCATATTCTATTAGTATATATATTTCCAAATTTTGAC 780
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 167 IlegIhIlleIleAlaGlyAlaLeuLeuIlegIhInPheLysGln-----Gly 181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 TATAACAATTAATTAATTTTCATTTGATGATATATATGATCTTACTTGATATGAAATGACA 840
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 LeuAspArgMeThrAlaCylSerAspAspLeuLysGlyPheLysAspIleIleThr 201
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 CGT-----TTTCAGTTACTCAACCATATACAGATTTAAATATCTTAT 885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 202 LysLysPheLysLysValPheProThrLeuSerLeuLysLysLysIleValAlaArgTyr 221
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 886 ACACATTTATCATTTCCATGATGATATGAAATTTTATGACTATATTAATCATTCATGCTGTGT 945
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 222 Ala-----TyrArgGly 225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 946 GGTTTAAATATGATATAACACCAATATACATAAACAACATATGATGAGCCTGTTTTCT 1005
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 226 GlyPheThrTrpLeuAsnAspArgPheLysGlyLysGlnIlegIyGlnGly--MetVal 244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1006 ATTGACATCAATTCGAGTATTCCTTATGAGATATATGATGAAATAATTCACATGATGTA 1065
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 245 PheAspValAsnSerLeuLysProAlaGlnMetLysSerArgLeuLeuPro----- 261
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1066 TACTTTTACGACACATTCACAGACCAACGTTAAATCCCT--ACTTTTATGATGATGAC 1122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 262 -----TyrGlyGlnProIleValPheGlnGlyLysTyrValIlyrAspGln 276
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1123 AATTTAT-----TTTTCATTTATATAGATTTGATA 1152
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 277 AspIlyrProLeuHisIleGlnHisIleArgCysGlnPheGlnLeu-----Lys 292
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1153 GATGATTTTAACGATGATTTTATTAATTAATTAATTAATTAACGATATTTAGCTCAATGATT 1212
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 293 GluGlyTyrIleProThrIleGlnIleLys--ArgSerArg----- 305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

QY 1213 GTAATAATACATATATATGATGATGATTAATATATACATTAAGTAAAG 1272
Db 306 -----PheTyrIysGlnValTyrLeuLysSerSerGlyGlnIleAlaasp 322
QY 1273 ATTCAGACATACAGGATATGATTCAGACATATACGATGATTCGTTGTATATAT 1332
Db 323 LeuTrrP---ValSerAsnValasp---LeuGlnIleMetLysGlnIleTyrAspLeuTyr 340
QY 1333 GAATGATGATAC-----TTTCATGCACGCGATATTTTTCATAAACTAT 1377
Db 341 AsnValGlnTrrIleSerGlyLeuLysPheLysAlaThrThrGlyLeuPheLysAspPhe 360
QY 1378 TTT-----ATTAAACA-----CAAGTAAGTTAAACAAACAAATCAAT 1416
Db 361 IleAspLysTrrPThrIleLysThrThrSerGlnGlyAlaIleLys----- 376
QY 1417 ATGACATACACCTTACACATTCATTCATGATATATACGACACCCATCTCAAT 1476
Db 376 ----- 376
QY 1477 GAGGAGTTATGTTATCTAAAGCGTTTAAATGATTAATAGGC----- 1521
Db 377 -----GlnLeuAlaLysLeuMetLeuAsnSerLeuTyrGlyLysPheAlaSerAsn 393
QY 1522 -----ATACCTGCATTAACGTTACATCTTAACCTTA--TTCCGTTTA 1560
Db 394 ProAspValThrGlyLysValProTyrLeuLysGlnAsnGlyAlaLeuGlyPheArgLeu 413
QY 1561 GATGATACATACATACATATACATATACGTTTACAAAACGTCAGCATATATA 1620
Db 414 -----GlyGlnGlnIleuThrLysAspProVal 422
QY 1621 TTA-----TTCTCTCATTTGTGCATGCATCGTTCTGATTAATGATTCCTTTC 1674
Db 423 TyrThrProMetGlyValPheIleThrAlaThrAlaArgTyrThrIleThrAlaIa 442
QY 1675 CAATCTTAACGGAAGTGAATGACGACATTTATTTATGCTGCTGATGATTCG 1734
Db 443 Gln-----AlaCysPheAspArgIleIleTyrCysAspThrSpsSerIle 457
QY 1735 TATATGAATCCGTTGTTAAACCCCTTATGAACCCAGTTTATTCGACCCGATAGCCTTA 1794
Db 458 HisLeuThrGlyThrGlnIleProAspValIleLysAspIleValAspProLysLysLeu 477
QY 1795 GGTAAATGCGATTTGAAAACGACGATGATGAATGTTTGTACTGATCATAGAAGA 1854
Db 478 GlyTyrTrpAlaHisGlnSer---ThrPheLysArgAlaLysTyrLeuArgGlnLysThr 496
QY 1855 TATGCATATGACGATGGAATGGAAGATTAAATGCTTCTGCTGATWACCGAAAAACGCC 1914
Db 497 TyrIleGlnAspIle----- 501
QY 1915 TTTCATACAAAGCGTCGATTTTGAACCTTTGTACGTAACATTTCTTGAACGCGCANT 1974
Db 502 -----TyrMetLysGln-----ValAspGlyLysLeu 510
QY 1975 ATGGAACATATAAAGTATCTATATGAGCAAGTACATATCGATATACGCTTAA 2034
Db 511 ValGlnIleSerProAspAspTyrThr-----ThrIle-----Lys 522
QY 2035 ACTGAATTCATGCTGTAATGATATATGATGATATTTTACTGATGATCACTTAATAGAA 2094
Db 523 PheSerValLysCysAlaGlyMetThrAspLys-----IleLys 535
QY 2095 CGTGAATTTATTTAAAGCGCTAGAGAAATTC 2130
Db 536 LysGlnValThrPheGlnAsnPheLysValGlyPhe 547

```

```

C:\Accession: S51275
R:\Martin, A.C.: Lopez, R.; Garcia, P.
submitted to the EMBL Data Library, January 1995
A:\Description: Nucleotide sequence and transcription of the left early region of Stre
A:\Reference number: S51271
A:\Accession: S51275
A:\Status: preliminary
A:\Molecule type: DNA
A:\Residues: 1568 <MAR>
A:\Cross-references: EMBL:Z47794; NID:92288892; PIDN:CAA87725.1; PID:9836593
C:\Genetics:
A:\Start codon: GTG
C:\Superfamily: phage PZA DNA-directed DNA polymerase

Alignment Scores:
Pred. No.: 2,07e-06 Length: 568
Score: 225.50 Matches: 151
Percent Similarity: 36.35% Conservative: 86
Best Local Similarity: 23.15% Mismatches: 258
Query Match: 5.54% Indels: 157
DB: 2 Caps: 32

US-09-727-892a-2 (1-2286) x S51275 (1-568)
QY 175 GAAGTATTTCCGAGTTTCGAATCTTTTATGACCATTTTATACGATATGGAAGACGT 234
Db 38 LysValAsnThrSerLeuGlnAspPheLeuLysSerLeuTyrLeuAspLeuAspLysThr 57
QY 235 GATACATACAAATCAAAACAGATATATATGATTCATGATTCGACATACGTAATATAATAC 294
Db 58 TyrThrGlnThr---GlyGlnAspGlnPheIleIlePhePheHisAsnLeu---LysPhe 75
QY 295 GATAATCATTTTCTTACTTAAAGACACATCGCTTATTTGTAATATATATACACGGGAAT 354
Db 76 AspGlySerPheLeuLeuSer-----PhePheLeuAsnAsnAspIleGlyCys 91
QY 355 ATATATTAAATCGCAGAGAAAGAAATGACACACATTAATAAGAAAGAGCTACTAT 414
Db 92 ThrTyrPhe----- 94
QY 415 TTAGCCAAAAATCAAAATGTAATTTTGAAGAAAGCGTTAAATCTTCATCAATTAAGAT 474
Db 95 -----IleAsnSpsMetGlyValTrrPtySerIleThrLeuGln 107
QY 475 TTAACAATGTTTAAATGCTTAAATTTATATATATGATGATGATGATGATGATGAT 534
Db 108 PheProAspPheThrLeuThrPheArg-----AspSerLeuLysIleLeuAsn 123
QY 535 ACATCAATTCGACACTTAGGT-----AAGAAATTCCTGATGGTGTTATTATTAACA 585
Db 124 PheSerIleAlaThrMetAlaGlyLeuPheLysMetProIleAlaLysGly---Thr 141
QY 586 GAATCACACCTTAACACGATTTTATATATACGATTTTGTATTAAGATATGATGAT 645
Db 142 ThrProLeuLys-----His 147
QY 646 GATACGAGCGCTATGACTATGCTGTAATGTTTGCAGAACTCACACCTGAACAACCT 705
Db 148 LysProGlnVal-----IleLysProGlnLysPhe 157
QY 706 ACATCAATTCATATGACGCTGATATATTA-----GATATGCGCATATTCATTAAGT 759
Db 158 AspTyrIleHisValAspValAlaIleLeuAlaArgGlyIlePheAlaMetTyrGly 177
QY 760 GATATATTTCGAATTTTGACATATACAAATTAACATTTTCATGATATATATGGAATCT 819
Db 178 GlnAsnPheThrLys-----TyrThrSerAlaSerGlnAlaLeuThrGlnPheLysArg 195
QY 820 TACTGGAATTAATGAATGACAGGTTTCAGTACTCAACCAATATCAAGATATTAATAA 879
Db 196 IlePheArgLysSerLysArgLysPheArgAspPhePheProIleLeuAspGlnLysVal 215
QY 880 TCTTATACATATATATCAATTTCCATGATATGAAATTTTATGACTATATTAATCATCTCAT 939

```



```

Db      447  LysLeuAlaLysAspHisnIleIleThrLysLys----- 458
      472  GATTTAACAATGTTTAAATGGTTTAAATTAAATTATGATTGATTATGAAACC 531
      458  ----- 458
      532  AATACATCAATTGCACATTAGGTAG-----AATTAATTGATGGTGGTTATTTA 582
      459  -----GlyLysPheProTyrLysPheValAsnLysAspAsnLeu 471
      583  ACAGAAATCACACTTAACACAGATTTTAAATTATAGCATTTTGAATAAGATATGATAG 642
      472  GltTyrValGlyLeuLeuProAspTyrGlyTyrLysPheProLysGlyGluMet 491
      643  AATGATAGACAGCCATGATGCTGCTGGAATGTTTGCAAAA-----CTCACACT 696
      492  -----IleThrLeuPheGluTyrPalaAlaMetTyrThrAsnLysTrpSerLeuArgLys 509
      697  GAACAATTAACATTAATTAATGACGTG-----ATTATATTAGGTATG 741
      510  GltThrIleIleTyrLeuGluLysAspIleLysAlaLeuTyrGlnLeuMetGluMet 529
      742  TGCCATATTCATTATAGTATGATATATTTCCAAATTTGACATTAACAATTAACATTTTA 801
      530  SerAsnAsnThrTyrSer-----ThrPheArg 538
      802  TTGAATATTATGAA-----TCT 819
      539  IleAsnIleThrArgValLysThrAlaSerAlaLeuAlaPheLeuValTyrArgThrIle 558
      820  TACTGATATATGAATGACACGCTTTGACTGACTGACCAATATCAATATTAATAA 879
      559  PheLeuProAsnGluValGluGluAsnGluThrAsnSerProAsnAsnIle----- 576
      880  TCTTAACATTAATCATTTCCATGATGATGAATTTTATGAC----- 921
      577  -----LeuSerLeuPheAspLysGluGluLys 587
      922  -----TATATTAAATCA 933
      588  LeuThrProLysTyrPheLeuProLysLeuLysGlyArgLeuGluArgAlaValArgAla 607
      934  TTCTATCGTCGCTTAAATATGCTATACCAACCAATATACATAAACAACATTAATGATGAG 993
      608  AlaTyrPheGlyGlyArgAsnGluIlePheIleProIleLeuAsnIle----- 624
      994  CCTGTTTTCATTCATGACATTCATGAGTATCCTTATGTCATCATGATAAATAAT 1053
      625  -----PheSerPheAspPheAsnSerLeuTyrProThrAlaMet-----MetMet 639
      1054  CCAACATGTTTACTTTTACGACACTATTCAGAACCAACGTTATCCCT-----ACT 1107
      640  ProMet-----ProValGlyIleProValHisThr 649
      1108  TTTTATGATGATGACATTTTTCATATATATTAAGATTGATTAAGATGATTATTAAGAT 1167
      650  PheCysLysAsnLeuAsn-----GluIlePheGly----- 659
      1168  GATTTTAATTAATAATTAATCACTGATTAAGTCGAATGATGTTAAATAATCTATAAT 1227
      660  -----PheValArgAlaLysIleIleThrProAlaIle----- 670
      1228  AATGATAATGATTACGTTAATATCATACAAATACAAATTAAGATGATTCAGACATTAGC 1287
      671  -----AsnIlePro 673
      1288  GGTATTGATTCATGATACATACGTTTAATTCGTTTGTATATGAAATGGAATACCTTT 1347
      674  ValLeuProCys---ArgValLysValAsnGly----- 683
      1348  CATGACGATATATTTTTCAAAC-----TATTTT----- 1380

```

```

Db      684  --ValGlnLysLeuIlePheProIleGlyGluTyrThrGlyTyrPheSerGluGlu 702
      1381  ATTAACACACAGGTAAGTAAATCAATATGACATACCTTACGATATACAC 1440
      703  LeuLysLeuAlaValGluTyrGlyTyrLysIleGluValLeuGluSerTyrValPheGlu 722
      1441  ATTACTGATCAT-----ATCACGACACACCA 1467
      723  LysArgAspAspProPheLysGluTyrIleGluHisPheAlaSerIleLysAspAsnThr 742
      1468  TACTCAATAGAGAGGTTATGTTATCTAAGTCTGTTTAAATGATATATGATACCT 1527
      743  LysGlySerLysLysGluMet---AlaLysLeuLeuAsnThrLeuTyrGlyArgThr 761
      1528  GCATTACCTGCATTTAACTTATTCCTTACATGATGATACGATGACATACATATAC 1587
      762  GlyLeuAsnAspSerAlaAlaGluIleLysMetLeuThrThrAsnGluLeuAspAsnIle 781
      1588  ATTAAGGTTACAAAACACGACGATATATATATTCCTCTACATTTGTC----- 1658
      782  -----GlnLeuThrAsnAsnValIleHisGluPheGluValAspAsp 796
      1639  -----ACATCAGTTCA 1650
      797  LysHisTyrValArgTyrAspLysLysProCysProValLeuCysAlaGlnSerGluLys 816
      1651  TTGATTAATTAATGTTCTTCCTTCCAAATCTTAACGGAAGTAATGACGATATTT 1710
      817  AsnTyrGluLeuLeu-----SerTyrLeu---AspGlyGluLysAspAspGlyPhe 832
      1711  AAT----- 1713
      833  IleIleAsnSerThrSerIleAlaAlaThrAlaSerTrpSerArgIleLeuMetTyr 852
      1714  -----TATGCGATACGATGATGTTTGTATGATGAATCCGTTGT 1752
      853  LysHisIleIleAsnSerAlaTyrThrAspThrAspSerIlePhe-----ValGlu 869
      1753  AAACCTTATTTGAACCCAGTTTATTCGACCCGATAGCCTTAAGTAATGATATGAA 1812
      870  LysProLeu-----AspSerAlaPheIleGlyGluGlyCysGlyLysPheLysAlaGlu 887
      1813  -----AACGAACGATGATGATGATGTTTGTACGATCTTAAGAAATGATATGAA 1866
      888  TyrAsnGlyGlnLeuIleLysArgAlaIlePheIleSerGlyLysLeuTyrLeuLeuAsp 907
      1867  GTGAATGGAAGATTAAATGCTTCTGCTGCTATACCGGAAAGCCCTTGATACAGC 1926
      908  PheGlyGlyLysLeuGlnIleLysCysLysGlyIleThrLysAsnLysAspAsnThrThr 927
      1927  GTGATTTTGAACCTTTGTACGTCGATGACAAATTCCTTGACGGTGCATATATGAACAAT 1986
      928  HisAsnLeuAspIleAsnAspPheGluAlaLeuTyrAsnGly-----GluSer 943
      1987  AAAATATCTATTAATGACAA-----GGTACAATATGATATAT 2025
      944  ArgValLeuPheGlnGluArgTyrPheLysGlyLeuGlyThrValThrValLys 963
      2026  CCGTCTAAACTGAATATGATGCTATGCTATATATATGATGAA 2067
      964  TyrGlnLysTyrAsnLeuIleSerLys-----TyrAspLys 975

```

RESULT 11
 D11606
 hypothetical protein PF0800C - malaria parasite (Plasmodium falciparum)
 C1Species: Plasmodium falciparum
 C1Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C1Accession: D11606
 R1Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
 Science 282, 1126-1132, 1998
 A1Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A1Reference number: A71600; M01D:99021743; PMID:9804551

A:Accession: D71606
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 11817 <GR>
A:Cross-references: GB:AE001418; GB:AE001362; NID:g3845275; PIDN:AAC71949.1; PID:g384528
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0800C

Alignment Scores:
Pred. No.: 5.82e-06 Length: 1817
Score: 218.00 Matches: 185
Percent Similarity: 35.54% Conservative: 159
Best Local Similarity: 19.11% Mismatches: 303
Query Match: 5.36% Indels: 322
DB: Gaps: 49

US-09-727-892a-2 (1-2286) x D71606 (1-1817)

```
QY 52 TTATCTGGGATATAGAA-----ACATTACCGTACAAATAAGTTAACGGA 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 LeuLeuTyAspLysGluLysLeuPheLeuPheThrTyAlaTyAspLysIleGlnThr 786

QY 97 -----CGAAAAAACCAACCAATATAAAAACGTTACTTATTCGTAGCAATTGGT 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 TyrThrTyArgLysGluLeuLysSerLysTyLysIleSerThrIleVal----- 803

QY 148 TGGTTTAATGGTTAAGAAATGATGATGATATTCGAGATTTC--GATCGTTTAT 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 -----AspArgAsnIleLysMetPheLeuLysPheLeuLysAsnTyAsn 818

QY 205 GACCATTATATACGTATGTGAAAAAGCGTGATACAAATCAAAATCAACAGATATT 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 819 AsnAsnGluAsnThrTyVal-----AspAsnIleLysSerLysAsnIlePhe 835

QY 265 ATCATGTGACATTAACCTGTAATTAATACGATTAAT----- 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 836 HisLeuLeuLysMetLysAsnLysValThrAsnLysThrAsnThrHisLysAspIle 855

QY 301 -----CATTTTACTTAAGACACACACGCGTTATTT-- 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 TyrGlnPheIleHisSerIlePtyHisIleLysLeuLysAspGlnAsnLysGluHisSer 875

QY 334 -----GATAATATTACACGGGAATAATATTAAATCTGACAGAAAT-- 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 876 PheGlnAspLysPtyLysIleAsnAsnLeuTyGluLysLysValGlnHisAsn 895

QY 382 -----GAACACACATTAAAAATG-----AAAGAGCGTACTATT--TTAGCCAA 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 ThrMetThrHisHisIleIleLysMetGluAspLysLysLysIleHisLeuMetGlu 915

QY 424 AATCAAAATGTAATTTAGAAAAACGTGTAATCTTCATCACTATTAGTTTAACATG 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 AsnAsnAsnMetLeuLeuAsnAsnAsnMetSerLeuAsnSerLeuAsn-- 934

QY 484 TTTTAAATGCTTTAAATTAATATTATGATTAACCTTTGGAACCAATACATCAAT 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 935 -----AsnSerIleProLeuAsn-----AsnSerIleProLeuAsnAsnSerIle 949

QY 544 GCAACATTAGTAAGAAATTACTTGATGTGTTATTTAACACAATCACAACTTAACA 603
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 ----- 949

QY 604 GATTTATATTATACGATTTTGTAAAGATAATGATGATGATGAGCCATGAC 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 ProLeuAsnAsnSerIle-----ProLeuAsnAsnSerIleProLeuAsn 964

QY 664 TAAGCTGTGAATGTTTGCAAAACGTACACCTGACACAACTTTCATCATCAATCAAT 723
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 AsnSerIleProLeuAsnAsnSerIleSerLeuAsnSerCysLysIleSerTyAsnSer 984

QY 724 GTGATTATATAGTATGTCATATTCATTAAGTATATATATTCGAATTTTGAC-- 780
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 985 IleSerLeu-----TyrSerAsnLysAsnThrSerPheAsnHis 997

QY 781 ----TATACAAATTAACATTTTCATTGAAATATTATGAAATCTTACTGATAATGAATG 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 998 LeuTyAsnAsnIleTyAsnThrCysPheIleGlnAsnAsnTyLysIleSerAsnGln 1017

QY 838 ACA-----CGTTTCAGTTCTCCACCAATACCAATACATAACCAACTAATGATGACCT 879
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1018 ValGlnAsnTyLysAsnGluLysAsnThrAsnMetGluHisTyLysGlnLysLysLeu 1037

QY 880 ----TCTTATACATTAATCATTCATGATATGATGATGATTTATGACTATTAATCATTC 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1038 PheIleTyProIleTyThrLeuGluAspLysAsnTyPhe----- 1051

QY 937 TATCGTGGTGGTTAATATGTAATACCAACCAATACATAACCAACTAATGATGACCT 996
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1052 -----LeuAsnValValAsnAsnIlePhePheAsnLys----- 1062

QY 997 TGTTTYTCTATGACATCAATTCGAGTTATCCCTATGTGATGATGATGAAAAATTCGA 1056
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 -----AsnTyAsnAsnThrPhePheTyThrCysGlnIleAsnIleLeuSer 1078

QY 1057 ACATGCTTACTTTTACGACACCTATTCAGAACACAGCTAATCCCTACTTTTAGAT 1116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 LysGlyLeuTyThrPheIleAsnTyTyThr-----ThrLeuLeuIle 1092

QY 1117 GATGACAATATTTTCATTAATATTAAGATTGATTAAGATGATTTAAGCATGATTATTA 1176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1093 SerSerAsnTyLysAla-----GluGlnIleLysThrAspAsnLys 1107

QY 1177 ATTAATAAT----- 1185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 CysAsnIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 1127

QY 1186 ----AAATCAGCTGATATTACGTCGAATGTTGTAATTAATTAATATGATATGCTTC 1242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1128 AsnAsnAsnAsnAsnAsnAsnAsnTyAsnAsnAsnAsnTyAsnAsnValTy 1147

QY 1243 GTTAATATCAATACAAATACATTAAGAAATGATGATGACAGCATTAACG--GCT 1290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1148 ProLeuIleAsnHisPheThrThrPheTyGluMetValThrTyLeuLeuLysAsn 1167

QY 1291 ATTGATTGCATGCATATA----- 1308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 IleTyArgIleHisIleSerLysPhePheTyIlePheValAlaLeuSerLysPhePhe 1187

QY 1309 -----CGTGTTAATTCG----- 1320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1188 LeuMetAsnSerTyGlnGlnSerAsnThrAsnLysArgGluAsnSerIleHisMetGlu 1207

QY 1321 -----TTTGTTATATATGAATGGAATGGAATCTTTCATGACAGTGATATT 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1208 AsnValLeuTyThrLeuTyThrIleLeuTyThrLysGlyLysGluHisValLysSerIle 1227

QY 1363 ATTTT-----CAAACTATTTTATTAATAAACAAGATGTTAAAAAAACA 1410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 LeuTyAspLysSerAsnGlnAsnTyThrPheArgPheAsnGluAsnLysAspIleLysMet 1247

QY 1411 ATCAATATGACATCACTTACGACTATCAATT-----ACTGATATATCAAC 1458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1248 GluAsnThrAsnMetLeuTyAsnIleIleLeuAsnAsnPheSerThrLysPheAsp 1267

QY 1459 GAACAC-----CCATCTCAATGAGGAGGTATGTTATCTGAAGTCGTTTAAAT 1509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1268 GluPheMetThrLeuGlnLysAsnGluAsp-----AsnAsnLysMetIleIleAsp 1285

QY 1510 GCAATTATGCAATCACTGCATTAAGTTTAACTTAATTCGTTTACATGATAC 1569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1286 AsnIleAsnValAspAsnIleAsnAspLeuIleLysSerHisHisCysAspAsnAsn 1305

QY 1570 -----AATGAAGTATTAAT----- 1584
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1306 LysLysGluAspThrSerSerLeuHisAsnLysLeuTyAsnGlyLeuHisPheLeuIle 1325
```



```

Db 540 GlnTyrAspAsnSerLeu----- 545
QY 685 AACTCACACCTGACAACTTACATCATTCATATGACGCTGATTATATAGGTATGCG 744
Db 546 -----PhePheGlnAsnAspLeuIleGlnIlePheAla 557
QY 745 CATATTCATTATAGTGATATATTCACAAATTTGACTATACAAATTAACATTTTCATTG 804
Db 558 ---LleIleMetLeuValIleLeuProLeuTyrGlnAsnArgLysVal----- 572
QY 805 AATATATGAACTCTTCTGGAATATGAAATGACAGCTTTTCGACTCAACCAATAT 864
Db 573 -----TyrLeuAspSerHisVal-----IleAsnLeuPheHisGlnIle 585
QY 865 CAAGATATTAATATCTTATACACATTAATTCATTCATGATGAATTTTATGACTAT 924
Db 586 TTrpAsnIleLysValGlnTyrAsp-----GluProLeuLysLysLeuProTyr 601
QY 925 ATTAAATCATCTCTGCTGGTTAATATATGATATACCAACCAATCATTAACCAACTA 984
Db 602 SerLysAsnGlnTyr-----AsnAspValGlnAsnAspTyrIleHisLysAla 617
QY 985 ATTGATGAGCCTGTGTTTTCTP----- 1011
Db 618 AspAspGlnIleCysIleAsnGlnLysValTyrAspGlnAspThrAsnLysTyrIleAsp 637
QY 1012 ATCAATTCGAGTTTCTCTTAT-----GTGATGTATCATGCAAAATTCACACATGG 1062
Db 638 ThrSerProAsnPheSerTyrAsnHisAsnIleGlnAsnHis----- 651
QY 1063 TTATACCTTTACGACACTATTCAGAACCAAGTTATCCCTACTTCTTTAGAT----- 1116
Db 652 ---TyrMetValGlnGlnHisSerGlnAspLysLys---ProTyrTyrMetAsnLysIle 669
QY 1117 -----GATGCAATATATTTTCA----- 1134
Db 670 LysTyrIleLysLysAsnAspGlnPhePheGlnGlnHisMetLysMetTyrGlnSerMet 689
QY 1135 ---TTATATTAAGATTTGATTAAGATGTTTAACGATGATTTATTAATTAATTAATCA 1191
Db 690 LeuIleTyrAsnArgSerLysLys---SerAsnAspAsnAsnThrIleProValAsnAsn 708
QY 1192 CGTGTATATACGTCAAATGANTGTAAATACTATATATATGATTAATGATTCAGTT- 1245
Db 709 LysMetGlnLysSerLysGlnLysAsnLeuAspAspAspArgLysAsnHisValLysLeu 728
QY 1246 -----AATATCAATACAAATTAATTAATGATG 1272
Db 729 LeuPheGlnIleAspAspLysAsnValSerAsnAsnAsnAsnThrAsnAsnIleAsnSer 748
QY 1273 ATTCAGACATTTACGGATATTGATGATGATGATGATGATGATGATGATGATGATGAT 1332
Db 749 IleAsnAsnIleAsnAsnIleAsnValCysAsn---AsnAsnAsnThrTyrValTyrAsn 767
QY 1333 GAATGTGAATACCTTCATGACCGGTGATATATTTTCAAAACTTTTATTAACACCA 1392
Db 768 LysValGlnPheLysSerIleCysAspLeuLysLysLys---TyrSerCysLys----- 784
QY 1393 GGTAGTTTAAAAAACAATCATATATGATCACCCTTAC-----GACTATCAC 1440
Db 785 ---GlnLeuLysAspGlnGlnLysGlnLysSerAsnProLeuLeuArgLysGlnGlnArg 803
QY 1441 ATTCTGATGATATCAACGAACACCACTACTCAATGAGGATTATGTTATCTAAAGTC 1500
Db 804 LeuHisSerAsp----- 807
QY 1501 GTTTTAATGATATATATGCGATACCTGCAATTCACATTTAATCTATTCGCTTAA 1560
Db 808 -----AsnIleAsnPheAsnLysLeu 814
QY 1561 GATGATTAACAATGAATATACATATATACGTTTACGTTCAAAAAACACTGAACGTAAATATA 1620

```

```

Db 815 AspAsnAsnAspAsp-----AsnIleGlnSerAspLeuLeuAspAsnAspLysArgTyr 832
QY 1621 TTATTCCTACATTTGTCACATCCATCCATTCATATTAATGATTTATGCTTCCTTCCAAATAC 1680
Db 833 TyrAsnLysGlnSerValArgLysGlnSerIleTyrAsn----- 845
QY 1681 TTACCGGAAGTAATATGACGACCAATTTTATTTATGCGATACGATAGTTGTATATG 1740
Db 846 ---GluAspArgIleCysGlnLysTyr-----LysLysSerLysLeuGlnTyr 860
QY 1741 AATCCGTTGTAAACCTTATATGACCCAGTTTATTCGACCCGATTAACCTTA----- 1794
Db 861 AsnGlnTyrLeuGlnLysPheMetAsnIleGlnIleArgAsnProPheIleIleLeuTyr 880
QY 1795 -----GCTAAATGCGATATTGAAAACGAACAGATATGATAGATGTTT 1836
Db 881 GlnLeuIlePheLysAsnPheArgMetAsnLeuAsnIleArgLysIleAsnAspMetTyr 900
QY 1837 GTA-----CTGAATCATTAAGATATATGCTATGAGTGATGAAAGATTAAATTT 1887
Db 901 ValGlnIleGlnLysIleHisIleLysMetAlaAspLysIlePheSerIleTyr----- 919
QY 1888 GCTTCTGCTGTATACCGAAAACGCCCTTGATATACAGCGTCGATTTTGAACCTTTGTA 1947
Db 920 -----ThrAsnLysSerThrValAspSerSerLysIleGlnHisIlePheTyr 935
QY 1948 CGTGACAAATCTCTTGACGCTGCCATTTATGAAAACAATTAAGTATCATATATAGCAA 2007
Db 936 PheGlnPhePheTyrAspIleIleIleArgIlePheAsnGlnLeuTyrTyrLeuGlnTyr 955
QY 2008 GGTACATATCGATATATATCCTGCTAAACGTAATGTATGTGATATATATGATGAA 2067
Db 956 SerAsnIle-----ProArgLysThrSer-----TyrGlnLys 966
QY 2068 TATTTT---ACTGATGACTTAATATGAACCGTGAATTTATTAATAAGCGTTAAGAA 2124
Db 967 TyrIleGlnArgLysLysTyrIleHisValLysLysIleLeuIleLeuLeuAlaSerLys 986
QY 2125 AATTTGACCATATGATGATGATATT 2154
Db 987 AspPheGlnTyrPheGlnLeuGlnVal 996

RESULT 13
E90097
hypothetical protein orf1019 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: E90097
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; M0ID:11323671; PMID:11323671
A:Accession: E90097
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1019 <DOO>
A:Cross-references: GB:AF165818; NID:g13794533; PIDN:AAK39908.1; GSPDB:GN00150
C:Genetics:
A:Gene: orf1019
A:Map position: 1
A:Gene: nucleomorph
C:Keywords: nucleomorph

Alignment Scores:
Pred. No.: 1,1e-05 Length: 1019
Score: 213.50 Matches: 186
Percent Similarity: 35.79% Conservative: 110
Best local Similarity: 22.49% Mismatches: 275
Query Match: 5.25% Indels: 256
DB: 2 Gaps: 47

US-09-727-892a-2 (1-2286) x E90097 (1-1019)

```

QY 61 GATATGAACATTAGGCTACATTAAGTTAACGGAGAAAAAACCAACCAATATATA 120
 ||||| : : : : : |||||
 Db 322 AsplIleuLysIleArgPheHISLysIleAsnIleArgLys ----- 335
 QY 121 AACGTTCTATCTGTAGCAATTGGTGTGTTAATGCTTAACAATTGATGTTGAAGA 180
 ||| : : : : : |||
 Db 336 -----IleTyrLysLeuAsnIle-----PhePheProPheAsnIleLeuValTyrPhe 351
 QY 181 TTTCGAGTTTCGATACCTTTTATGACGCAATTTTATACGATGTGAAAAGCGATGATACA 240
 ||||| : : : : : |||||
 Db 352 ThrValSerPheLysHISPhleu ----- 359
 QY 241 ATCAACAACCAACCAACAGATATATATCATGATTCGCATTAAGCTATATAATACGTAAT 300
 ||||| : : : : : |||||
 Db 360 -----SerLysLeuAsnSerIlePheTyrLeuHISgluPheSerGln ----- 373
 QY 301 CATTTTCTACTTAAGACACACATCGCTTATTTGTAATATATTAACGCGCAAAATATATAT 360
 ||||| : : : : : |||||
 Db 374 -----LeuLysCysLeuAsnLysIleIle-----IleThrAsnAsnAsnIleTyr 388
 QY 361 TTTAAATCTCCAGAGAAAAATGACACACACTTAAATATGAGAGGCTACTATTTTAGCC 420
 ||||| : : : : : |||||
 Db 389 -----AspGlu -----LysLeuLysThrIleThrAsnLeu --- 398
 QY 421 AAAAATCAAAATGTAATTTTAGAAAAACGTTAAATCTTCATCAATTAGATTAGACA 480
 : : : : : ||||| : : : : : |||||
 Db 399 -----MetPheLeuGluLysAsnLeuLysSerSerLeuSerMetSerLeuVal 414
 QY 481 -----ATGTTTTTA 489
 415 AsnAlaIleTyrSerLeuAsnAsnIlePheGluLysAsnGluLeuSerLysHISile : : :
 490 AATGGTTTAAATTTAATATATAT -----GATAC 519
 ||||| : : : : : |||||
 Db 435 TyrGluPheLeuPheAsnLeuIleIleTyrProProTyrValLysPheTyrGluAsn 454
 QY 520 TTTATGAAAAACAATACATCAATTCGACAACTTAGTAAGAAATTAAGTGGTGGTAT 579
 ||| : : : : : |||
 Db 455 IleMet -----IleLysTyrGluAsnPhe 462
 QY 580 TTTACAGAAATCACAACTTAACACAGAT -----TTTAATATATACGATTTTGTATAA 630
 ||||| : : : : : |||||
 Db 463 LeuIleIleAspAsnIleGlnThrAlaLysLysLeuPheLysLeuThrIleThrGluLys 482
 QY 631 GATTAATGATATGAT -----GATAGTGAAGCCTATGACTAGCTGGAATGT 678
 : : : : : ||||| : : : : : |||||
 Db 483 AsnSerAspIleAsnPhleValIleLeuGluGlnAsnLysTyrGluTyrLysGluAsn 502
 QY 679 TTTCGAAAACTCACACCTGACACTTACATCATTCATTAAGACGTGATATATAGCT 738
 ||||| : : : : : |||||
 Db 503 LeuMetLysLeuSer -----LysLeuIleArgVal ---AsnGluLeuIle ----- 516
 QY 739 ATGTGCATATTCATATATAGTATATTTTCCA -----AATTTGACTAT ----- 783
 ||| : : : : : |||
 Db 517 TyrCysLysLysLysPheLysThrIleIleAsnSerIlePheAsnThrSphLeuThr 536
 QY 784 AACAAATTAACATTTTCATTTGAATATATGAAATCTTACTGTAATATGAAATGACACGT 843
 : : : : : ||||| : : : : : |||||
 Db 537 HISAsnLeuAlaLysSerLeuTyrValAlaLysLysTyrLysIleAsnIlePheThrGlu 556
 QY 844 -----TTTCAGTTTACTCAACCAATATCAAGATATTAATATCTATACATATATCAT 897
 ||||| : : : : : |||||
 Db 557 IleGluPheGluIleArgAsnGluAsn -----LysIleLysTyrIleAsnGluLys 573
 QY 898 TTCCATGATATGAATTTTATGACTATATTAATCATCTGATCGTGGTGTAAATATAG 957
 ||| : : : : : |||||
 Db 574 LysHISPhleValAsnPhleIleSerTyrLeuLys ---IleGluArgThrGluLeuLysGlu 592
 QY 958 -----TATAACCAAAATTAACATAAACAATAATTTGAT 990
 ||| : : : : : |||||
 Db 593 IleLeuSerMetLysLeuThrMetAspTyrLysIleIlePheLeuLysLeu ----- 610

QY 991 GAGCCTTGTTTCTATATGACATCAATTCGACTATTCCTTAGTGATGATATGAAAAA 1050
 ||| : : : : : |||||
 Db 611 GluGlnIlePhe -----IleLysIleLysGlnThrTyrAsnLysIleLysPheHISerSer 629
 QY 1051 AATCCACACATGGTTTACTTTTACGACACATTTTACAGAACCAACGCTATATCCCTACTTT 1110
 : : : : : ||||| : : : : : |||||
 Db 630 LeuAspSer -----TyrSerLysPheAlaArgGluAsnLeuSerIleThrTyr 645
 QY 1111 TTAGATGATGACAAATTTATTTTCATTA -----TATAAGATGATAAAGATCA 1137
 ||||| : : : : : |||||
 Db 646 SerIleHISLeuCysTyrPheIleLeuAsnAsnGlnLysPheAsnArgAsnPhleLysCys 665
 QY 1138 -----TATAAGATGATAAAGATCA 1158
 666 TyrIleAsnGlnPheAsnSerThrIleThrHISLysAsnTyrLysTyrGluMet ---Ile 684
 QY 1159 TTTAAGCATGATTTATTTAATTAATTAATCAACGCTGATTAACGCAATGATGTATAAA 1218
 ||| : : : : : ||||| : : : : : |||||
 Db 685 PheThrGluGlnIlePheIleSerTyrPheSerPheIleIleGluLeuPheIleAlaIle 704
 QY 1219 TACTATTAATATGATTAAGATTTACGTTAATATCAATACAAATACATTAAGATGATTCOA 1278
 ||| : : : : : |||||
 Db 705 AsnTyrLysLeuValThrLysLysPheLysMetAsnValLysTyrLeuLys ----- 721
 QY 1279 CACATTAACGGGATGATTGCATGCATATACGCTGTAATTCGTTGTTATATGATGAATGT 1338
 : : : : : ||||| : : : : : |||||
 Db 722 -----LysAsnAsnLysAsnLysThrTyrLysLeu 731
 QY 1339 GATTACTTTCATGCACGATATATATTTTCAAAACATTTTATTAACCAACAGATAG 1398
 ||| : : : : : ||||| : : : : : |||||
 Db 732 IleAsnPhe -----AspIleLeuPheGlnAsn -----LysAsnThrSerLeu 745
 QY 1399 TTA -----AAAACAAAATC -----AATATGACATGACCT -----TAC 1431
 : : : : : ||||| : : : : : |||||
 Db 746 IlePheTyrGlnAsnLysIleTyrIlePheAsnPhleMetSerArgIlePheGluLeuPhe 765
 QY 1432 GACTATCAATTAACATGATATATACACGACACCCATCACTCAATAATGAGAGATATGTA 1491
 : : : : : ||||| : : : : : |||||
 Db 766 AsnLeuTyrLeuIleArgAspSerGlnIleHISValIleLysAsnPhleIleAsnGln 785
 QY 1492 TCTAAGTCGTTTAAATAGATATATATGCACTACCTGCATTAAGCTTACATTTTAAGTCA 1551
 ||| : : : : : |||||
 Db 786 PheLysGluIle -----SerLeuLysTyrLeuPheLysLys 797
 QY 1552 TTCCCTTAGATGATTAACAAATGAACTATATACATATATGATGCTTACAAACACATGAA 1611
 ||| : : : : : ||||| : : : : : |||||
 Db 798 PheMetIleLeuThrLysAsnGluAsnIleAsnGluLysAsnSerTyr ---AsnTyrLys 816
 QY 1612 CGTATATATATCTCTCACTTTGTCACATCAGCTTCATGCTATATCTTATGTTGCTCT 1671
 : : : : : ||||| : : : : : |||||
 Db 817 GlnLeuPhePhePheValThrArgPheHISAsnLysSerGlnLeuAspLeuIleLeuLeu 836
 QY 1672 TTCCATATCTTAAGGAAGAAATGACGACAAATTTATTTATTCGATACGATAGAT 1731
 ||||| : : : : : |||||
 Db 837 PheGlnSerIleIleIleLeuAlaIle ----- 845
 QY 1732 TTGTAATGAATCCGTTGTTAAACCTTATTAACCCCACTTATTCGACCCGATAGCC 1791
 ||| : : : : : ||||| : : : : : |||||
 Db 846 PheLysLysLysArgIlePheArgPheLeuLeuAsn ----- 857
 QY 1792 TTAGTAATAGGATATGAAACAGACATAGATAGATGTTTGTATGATCATATAG 1851
 : : : : : ||||| : : : : : |||||
 Db 858 -----MetGlnHISGluIleValLysSerPhe ---LeuAsnPhleLys 870
 QY 1852 AATATATGCA -----TATGAGATGATGGAAGATTAATATGCTTGTGCTGATATACG 1905
 ||| : : : : : ||||| : : : : : |||||
 Db 871 PheTyrProIlePheTyrThrLysAsnAsnMetSerLysPheLysGluAsnLysLys 890
 QY 1906 AAAAAGCGCTTTGATACACAGCGCTGATTTTGAACCTTTGACGTGAACAATTTGTTGAC 1965
 ||| : : : : : ||||| : : : : : |||||
 Db 891 LysValLysPhe ---TyrGluIleAspPheIleThr ----- 901
 QY 1966 GGTCCCATTTATGAAACAAATAAAGTATCTAT -----AAT 2001

QY 943 -----GGTGGTTAAATATGCTATACACC----- 966
 Db 321 AsPTyrasnIleValIlePheLysGluValAsnAsnPhenheAsnAsnTyrAsnAlaPheIle 340
 QY 967 ---AAATACATA-----AACAACTAATGTAGAGCTTGT---TTTCTATT 1008
 Db 341 AspLysTyrIleAlaPheSerAsnAsnLeuAlaIleCysLysIleCysGlyGluSerIle 360
 QY 1009 GACATC-----AATTCGAGTTATTCCTTATGTATG 1038
 Db 361 AspNetPheAsnPhenValGluAlaAsnTyrIleGlnSerHisGlyTyrMetIleIleThr 380
 QY 1039 TATCATGAAAAATCCACAGTGTATACCTTTAGCAACCTATTCAGAACCAAGCTTA 1098
 Db 381 ThrHisLysAspAsn-----TlePheGlnTyrGluThrTyrGlyLysLeuThrAsn 397
 QY 1099 ATCCCTACTTTTATAGATGATGACAAATATTTTCATTAATAGATGATGAAGATGA 1158
 Db 398 AlaGluLeuPheLeuSerAsp-----TyrLeuSerIleTyr-----AspSerIle 412
 QY 1159 TTTAACCATGATTATTATTAATTAATCAACCTGATTAAGCTCAATGATTGTAAAA 1218
 Db 413 PheAsnThrAsnValMet-----AATTCGAGTTATTCCTTATGTATG 418
 QY 1219 TACTATAATATGATGATGATTAATCAATCAATACAAATACATTAAGATGATCA 1278
 Db 419 -----AspAspPhe-----AsnAsnThrAlaAlaGluLeuIle 429
 QY 1279 GACATTAACGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338
 Db 430 AspTyr-----MetIleHisIleAsn---AsnAsnArgLeuMetTyrGluGlu 444
 QY 1339 GAATCATTTTCATGCACCGATATATTTTCAAAACATATTATTAACCAACAGGT--- 1395
 Db 445 GluTyrLysHisGlu-----IleAspThrSerGlyLeu 455
 QY 1396 -----AAGTTAAAAACAAATCAATATGACATCACCTTACGACTATACACTTACT 1446
 Db 456 PhePheValArgLeuThrAsnAsnIlePheMetSer----- 467
 QY 1447 GATGATATCAACGAACACCCCTACTCAATCAATGAGAGGTTATGTTATCAAGTCGTTTA 1506
 Db 468 ---GluTyrAsnGluLysGluGlnPheArgGluGlnArgMetIleAsnIleMetIleVal 486
 QY 1507 -----AATGGATTATATGCGATCCTGCATTA 1533
 Db 487 IleIleIleThrLeuValLeuValAsnAsnPhenAsnGluLeuIleGlyIleValLysArg 506
 QY 1534 CGTTCACATTTTACTTA-----TTCGGTTAGATGATACATGACATGACATATACATATC 1587
 Db 507 LysAspMetPheLysArgIleAspPheLysGlyIleAsnGluLeuIleIleGluIle 526
 QY 1588 ATTAACGGTTAC-----AAAACACTGAACGTAATATATTA 1623
 Db 527 ValSerGluTyrIleValLysGlnGlyIleAspIleLysAsnLeuAsnIleProValIle 546
 QY 1624 TTCCTACATTTGTC-----ACATCACGTTGATGATTAATGATTTGGTT 1668
 Db 547 IleAsnThrTyrIleLysIleLeuThrProGluLeuLysSerHisTyrAspIleLeuVal 566
 QY 1669 CCTTCCAAATAC-----TTAACGGAAGTGAATGACGACACATTTTATT 1713
 Db 567 LeuArgPheTyrAsnHisIleAspIleLeuThrMetGluGluIle-----IleIle 583
 QY 1714 TATTGCGATGATGATGATTTGTATATGAATCCGTTGTAACCTTATTAAGACCCAGT 1773
 Db 584 Tyr-----AspPheProMetAlaAsnAlaValLysAsn-----ThrHisSer 597
 QY 1774 TTATTCGAC-----CCGATAGCCTTAGTAAATGGGATATTGAAAC 1815
 Db 598 LeuLeuAspLeuTyrThrAspLysIleProTyrValLeu----- 610
 QY 1816 GAACAGATAGATAGATGTTGTACTG---AATCATAGAATATGCAATATGAAGTGAAT 1872

Db 611 ---GluLeuAspAsnIleLeuThrTyrGluAsnAsnIleAsnTyrAlaIleLysAspLys 629
 QY 1873 GGAAGAATTAATATGCTTGTCTGTGATACCAAAAAC-----GCCTTT 1917
 Db 630 IleAsnIleSerIleAsnSerIleAsnIle---LysAsnTyrLysGluPheThrSerSer 648
 QY 1918 GATACAAACGTCGATTTTGAACCTTTGACGTGAACATTTCTTGACGCGCC----- 1971
 Db 649 AspIleAsnIleGluLeuLysSerLeuIleSerGluIleLysPheGluTyrThrTyrLys 668
 QY 1972 -----ATTATGAAAC 1983
 Db 669 AsnThrLeuValArgValLeuLysGluIleAspAsnAspIlePheTyrIleAspAsnSer 688
 QY 1984 AATAAAGATCTATATATGAGCAAGGTACATATGATATATCCGTAAACTGAAAT 2043
 Db 689 HisLysTyrPhePheAsnThrAsnGluIleIleAsnGluSerProPheLys----- 705
 QY 2044 GTATGCTGATATGATATGATGAA-----TATTTACTGATGAACTT----- 2085
 Db 706 IleIleGlyAsnGluTyrPheLysLeuLeuTyrPheSerAspProLeuProPheGlu 725
 QY 2086 -----AATATGAACGTCGAATTTATTAATAAGACGCTGAGAAAAATTGACACATAGT 2139
 Db 726 AsnIleAsnLysLysHisLeuThrIleLeuTyrAspGlyIleAsnIlePheLeuAsnIle 745
 QY 2140 CAATTTGATGATATCTCTTATATGAA 2166
 Db 746 TyrPheProSerTyrThrPheValAsp 754

Search completed: January 8, 2003, 17:02:43
 Job time : 88 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame-plus-n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 21 Seconds
(without alignments)
9029.989 Million cell updates/sec

Title: US-09-727-892a-2
Perfect score: 4070
Sequence: 1 atggattactagaatgcac.....taaaaaaggaactgttaa 2286

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+np.model -DEV=xlh
-O=/cgn2_1/USPRO.spool/US09727892/rnat_06012003_151200_9288/app-query.fasta-1.2439
-DB=SwissProt-40 -QFMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US09727892.ecgn_1.1.15-etunal_06012003_151200_9288 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT-120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPO=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt-40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241.5	5.9	575	DPOL_BPPH2	P03680 bacterioph
2	234.5	5.8	572	DPOL_BPPH2	P18894 bacterioph
3	228.5	5.6	572	DPOL_BPPH2	P06950 bacterioph
4	225.5	5.5	568	DPOL_BPPH2	P37989 bacterioph
5	221.5	5.4	1021	DPOL_BPPH2	P35537 bacterioph
6	220	5.4	572	DPOL_BPPH2	P03682 bacterioph
7	208.5	5.1	1162	BXEN_CLOBO	Q06366 clostrid
8	208	5.1	797	DPOM_AGABT	P30322 agarcus bi
9	207.5	5.1	2136	YCF2_MARPO	P09979 marantia
10	207	5.1	1024	RPOB_PLAFA	P21421 plasmodiu
11	200	4.9	970	YK87_BUCAT	P57189 buchnera ap
12	198	4.9	1162	BXEN_CLOBO	P46082 clostrid
13	192.5	4.7	1197	DPOM_PODAN	P07625 podospo a
14	185.5	4.6	2339	RPCI_PLAFA	P16134 plasmodiu
15	183.5	4.5	2710	TOXA_CLODI	Q08372 plasmodiu
16	183	4.5	3135	S230_PLAFO	P10479 bacterioph
17	182	4.5	553	DPOL_BPPH2	P10479 bacterioph
18	178.5	4.4	760	EFV2_MSEPV	Q9YW29 melanoplus

19	175.5	4.3	964	DPOL_CBEV	P30319 christoneu
20	174.5	4.3	1024	Y075_MYGE	P47321 mycoplasma
21	173.5	4.3	986	EP1B_STAEP	P30191 staphylococ
22	172	4.2	995	DPOL_KLUFA	P09804 kluyveromyc
23	171	4.2	1956	ATX1_PLAFA	Q04956 plasmodiu
24	168.5	4.1	1481	RPOD_ODOST	P49468 odontella s
25	167.5	4.1	1630	MSPI_PLARK	P04932 plasmodiu
26	167.5	4.1	1639	MSPI_PLARK	P09933 plasmodiu
27	167	4.1	874	SVY_UREPA	Q09444 ureaplasma
28	166.5	4.1	1250	BXE_CLOBO	Q00496 clostrid
29	165.5	4.1	647	NTPI_CBEV	Q09179 christoneu
30	165.5	4.1	973	HYPA_CLOPE	Q46205 clostrid
31	165	4.1	575	RPOB_PLAFA	P21422 plasmodiu
32	165	4.1	876	RPOB_NPVAC	P41452 autographa
33	165	4.1	1196	BXEN_CLOBO	P46081 clostrid
34	163.5	4.0	780	MUS2_BORBU	Q51125 borrelia bu
35	163	4.0	1169	EX5B_BORBU	Q51578 borrelia bu
36	162.5	4.0	667	Y366_MYGE	P47606 mycoplasma
37	162	4.0	994	DPOL_KLUFA	P05468 kluyveromyc
38	161.5	4.0	2366	TOXA_CLODI	P18177 clostrid
39	160.5	3.9	817	RPOD_ASTLO	P58132 ascaris lon
40	159.5	3.9	648	NTPI_MSEPV	P29814 amastax moe
41	159	3.9	647	NTPI_MSEPV	Q9YW39 melanoplus
42	159	3.9	2167	BEM2_YEAST	P39960 saccharomyc
43	158.5	3.9	993	NISB_LACLA	P20103 lactococcus
44	158	3.9	717	NTPI_MSEPV	Q9YW06 melanoplus
45	158	3.9	1018	YCI4_METUA	Q58611 methanococc

ALIGNMENTS

RESULT 1
ID DPOL_BPPH2 STANDARD: PRT: 575 AA.
AC P03680:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (Early protein GP2).
GN 2 Bacteriophage phi-29.
OS Bacteriophage phi-29.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC phi-29-like viruses.
OX NCBI_TaxID=10756;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82262795; PubMed=6809534;
RA Yoshikawa H., Ito J.,
RT "Nucleotide sequence of the major early region of bacteriophage phi
RT 29."
RL Gene 17:323-335(1982).
RN [2]
RP SEQUENCE OF 1-85 FROM N.A.
RX MEDLINE=83064518; PubMed=6292852;
RA Escarimis C., Salas M.,
RT "Nucleotide sequence of the early genes 3 and 4 of bacteriophage phi
RT 29."
RL Nucleic Acids Res. 10:5785-5798(1982).
RN [3]
RP SEQUENCE FROM N.A. (TEMPERATURE SENSITIVE MUTANT TS2(24)).
RX MEDLINE=90370456; PubMed=2118623;
RA Bianco M.A., Bianco L., Pares E., Salas M., Bernad A.,
RT "Structural and functional analysis of temperature-sensitive mutants
RT of the phage phi 29 DNA polymerase."
RL Nucleic Acids Res. 18:4763-4770(1990).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=96123630; PubMed=8594366;
RA Lazaro J.M., Bianco L., Salas M.,
RT "Purification of Bacteriophage phi 29 DNA polymerase."
RL Meth. Enzymol. 262:42-49(1995).
RN [5]
RP MUTAGENESIS.

```

RX MEDLINE=96123647; PubMed=8594354;
RA Blanco L., Salas M.;
RT "Mutational analysis of bacteriophage phi 29 DNA polymerase.";
RL Meth. Enzymol. 262:283-294(1995).
RN [6]
RP MOTAGENESIS OF SER-122 AND LEU-123.
RX MEDLINE=99003248; PubMed=9786901;
RA de Vega M., Blanco L., Salas M.;
RT "phi29 DNA polymerase residue Ser122, a single-stranded DNA ligand for
3'-5' exonucleolysis, is required to interact with the terminal
RT protein.";
RL J. Biol. Chem. 273:28966-28977(1998).
CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMAIC ACTIVITIES:
CC DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOLYTIC ACTIVITY THAT
CC DEGRADES SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC -----
CC DR EMBL: V01155; CAA24480.1; -
CC DR EMBL: X53371; CAA57451.1; -
CC DR PIR: A04282; EERP29.
CC DR Interpro: IPR002064; DNA_pol_B.
CC DR Interpro: IPR004868; DNA_pol_B_2.
CC DR Pfam: PF03175; DNA_pol_B_2; 1.
CC DR PRINTS: PR00106; DNAPOLB.
CC DR SMART: SMO0486; POLB; 1.
CC DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
CC
CC Transferrase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Hydrolase; Exonuclease; Early protein.
CC
CC SITE 122 122 CRITICAL FOR 3'-5' EXONUCLEOLYSIS.
CC FT VARIANT 176 176 A -> R (IN MUTANT TS2(24)).
CC FT VARIANT 355 355 A -> V (IN MUTANT TS2(24)).
CC
CC SO SEQUENCE 575 AA; 66714 MW; 856EBB604A7E266 CRC64;

```

Alignment Scores:

Pred. No.:	1.79e-07	Length:	575
Score:	241.50	Matches:	151
Percent Similarity:	35.57%	Conservative:	114
Best Local Similarity:	20.27%	Mismatches:	239
Query Match:	5.93%	Indels:	241
DB:	1	Gaps:	40

US-09-727-892a-2 (1-2286) x DPOL_BPPH2 (1-575)

```

QY 31 AAACATGACGTCGATGATTTTATCTGATATGAAACATTAAGCGATATAAGTT 90
   |||||  |||  ::  |||  |||||
Db 2 LysHisMetProArgLysMetLysSerCysAspPheGluThr----- 15
QY 91 AACGACGAAAAAACAACCAATATATAAAGTTACTATTCTGTAGCAATTGGTTGG 150
   |||||  |||||  ::|||  ::|||
Db 16 -----ThrThrLysValGluAspCysArgVal-----Trp 25
QY 151 TTTAATGTTTGAATGATGATGATTTTCGAGTTTCGAA-----TCTTTTAT 204
   |||||  |||||  ::|||  ::|||
Db 26 AlaItyrGlyTyr---MetHisIleGluLysPheSerGluItyrLysIleGlyAsnSerLeu 44
QY 205 GACGATTTTATACGTATGTGTAAGAAAGCGTATACATACAAATAACAAACAGATATT 264
   |||  |||  ::|||  ::|||
Db 45 AspGluPheMetAlaItyrVal-----LeuLysValGluAlaAspLeu 58
QY 265 ATCATGATTCGACATACGTATATAATACGATATATCTTTTTCCTTAAGACACCATG 324
   |||||  |||||  |||||  |||||
Db 59 Tyrrpe-----HisAsnLeu---LysPheAspGlyAlaPheIleIle----- 71

```

```

QY 325 CCGTATTTTGGATATATATACACGCGAAATATATATTTAAATCTGCAGAAATAATGAA 384
   |||  |||  |||  |||
Db 72 -----AsnIrrLeuGluArgAsnGlyPheLysTrpSerAlaAspGly----- 85
QY 385 CACACATTTAAATAATGAAGAGCGTCTACTTTTAGCCAAAATCAAAATGTATTTAGAA 444
   |||  |||  |||  |||
Db 86 -----LeuProAsnThrTyrrAsnThrIleIle--- 94
QY 445 AAACGCGTTAAATCTTCATCAATTTTATGATTAACAAGTTTAAAGTTTAAATTTT 504
   |||||  |||||  |||||  |||||
Db 95 SerArgMetGlyGluIrrTyrrMetLysPheIleGlySerLeuGlyItyrLysGlyLysArg--- 113
QY 505 AATATATTTGATTAATCTTATGAAAACCAATTCATCAATTCGACATTAAGTAAGAAATTA 564
   |||  |||  |||  |||
Db 114 -----LysIleHisThrValIleItyrAspSerLeuLysLysLeu 126
QY 565 CTTGATGGGTGTTATTTAACAGATACAGATTAACAAGATTTTAAATTAATAGCATTTT 624
   |||||  |||||  |||||  |||||
Db 127 -----ProPheProValLysLysIleAlaLysAspPheLysLeuThrValLeu 142
QY 625 GATAAAGATTAATGATATGATAGATAGTGAAGCCATGACTATGCTGCAAAATGTTTGA 684
   |||  |||  |||  |||
Db 143 LysGlyAspIleAspTyrHisLysGluArgProValGlyTyr----- 156
QY 685 AAACACACACCGTGAACACACTTACATACATTCATATGACGCG---ATTATATTAGTATG 741
   |||||  |||||  |||||  |||||
Db 157 LysIleIrrProGluGluItyrAlaItyrLysLysAsnIrrGluIleIleAlaGluAla 176
QY 742 TGCCATATTCATTAATAGTATATATTCACAAATTTGACATTAACAATTAACATTTTCA 801
   |||  |||  |||  |||
Db 177 LeuLeuIrrGluPheLysGlu-----GlyLeuAspArgMetThrAlaGly 192
QY 802 TTGATATATTATGCAATCTTACTTGAATATATGAAATGACACGT-----TTT 846
   |||  |||  |||  |||
Db 192 SerArgSerLeuLysGlyPheLysAspIleIleThrThrLysLysPheLysValPhe 211
QY 847 CAGTTACTACCAATATCAAGATATTAATAATCTTATACACATTCATCTTCATGATGAT 906
   |||||  |||||  |||||  |||||
Db 212 ProThrLeuSerLeuGluLysLeuAspLysGluValArgIrrAla----- 225
QY 907 ATGATTTTATTGACTATATTAATATCATCTTCATGCTGCTGCTTAAATATGATATACACC 966
   |||||  |||||  |||||  |||||
Db 226 -----TyrArgGlyGlyPheThrIrrPheLysAsnAsp 235
QY 967 AAATACATTAACAACATAATGATGACCTGTTTCTATGATGACATCAATGAGATTAT 1026
   |||||  |||||  |||||  |||||
Db 236 ArgPheLysGluLysGluIrrGlyGlyGly---MetValPheAspValAsnSerLeuTyr 254
QY 1027 CCTATGCTGATGATCATGAAAAAATTCACACATGCTTACTTTAGCAACACTATTCAT 1086
   |||  |||||  |||||  |||||
Db 255 ProAlaGluMetLysSerArgLeuLeuPro-----TyrGly 266
QY 1087 GAACCAAGCTTAATCCCT---ACCTTTTATGATGATGACAAATTAAT----- 1128
   |||||  |||||  |||||  |||||
Db 267 GluProIrrLeuAlaPheGluGlyLysTyrValIrrProLysPheLysIrrGlu 286
QY 1129 -----TTTTCATTAATTAAGATTTGATTAAGATGATTTTAAGATGATTTTA 1173
   |||||  |||||  |||||  |||||
Db 287 HisIrrArgCysGluPheGluLeu-----LysGluGlyTyrIrrProIrrIle 302
QY 1174 TTAATTAATAATTAATACACGCTGATTTACGTCATATGATTTGTAATAATCTTAATATGAT 1233
   |||||  |||||  |||||  |||||
Db 303 GluIrrLys---ArgSerArg-----PheTyrLysGly 312
QY 1234 AATGATTAAGTTAATATACATATACAAATATACATTAAGATGATTTACAGCATTTAGGGTATT 1293
   |||||  |||||  |||||  |||||
Db 313 AsnGluIrrLeuLysSerSerGlyGlyIrrIrrAlaAspPheIrrP---LeuSerArgVal 331
QY 1294 GATTCGATGCAATACGCTGTTAATTCGTTGTTATATATATGAATGATTC----- 1344
   |||  |||  |||  |||
Db 332 Asp---LeuIrrLeuMetLysGluIrrHisTyrAspLeuLysArgValGluIrrIrrSerGly 350

```

QY	1345	-----TTTCATCGACCGTGTATATTATTTTCAAAACTAT-----TTTAT	1383
Db	351	LeuLysSheLysAlaThrThrGlyLeuPheLysAspPheIleAspLysTyrThrTyrIle	370
QY	1384	AAAAACA-----CAAGCTAAGTATMAAAACAAATATCATATGACATCACTTACGACTAT	1437
Db	371	LysThrThrSerGlnGlyAlaIleLys-----	379
QY	1438	CACATTTCGTATGATATATCAACGACGACCCCATCTACTCAAAATGAGAGGTTATGTTATCTAA	1497
Db	380	-----GlnLeuAlaLys 383	
QY	1498	GTCGTTTAAATGATATATATGCG-----ATA 1524	
Db	384	LeuMetLeuLysSerLeuLysGlyLysPheAlaSerAsnProAspValThrGlyLysVal	403
QY	1525	CTTCGATTACGTTACGTTTAACTTA-----TTCCGTTTGATGATATACATGAACTATAC	1581
Db	404	ProTyrLeuLysGlnGlnGlyAlaLeuGlyPheArgLeu-----	416
QY	1582	AATATCATTAACGGTATGCAAAAACGTAACGCAATATATTA-----TTCTCTACATTT	1633
Db	417	-----GlyGlnGlnGlnThrLysAspProValTyrThrProMetGlyAlaPhe	432
QY	1636	GTCACATCAAGTTTCAATGATATTAACCTATTGGTTCCTTCCATACTTAAAGGAAAGTGA	1699
Db	433	IleThrAlaThrAlaArgTyrThrThrIleThrAlaAlaGln-----Ala 447	
QY	1696	ATTGACACACATTTATTTATTTGCGAATGACTGATGTTGATATGATGCAATCGTGTGTTAA	1755
Db	448	CysTyrAspArgIleIleTyrCysAspThrAspSerIleHisLeuThrGlyThrGluIle	467
QY	1756	CCCTTATTTGAACCCAGCTTATTTGACACCGCATAGCCTTAGATGAATGAGATATGAAMC	1815
Db	468	ProAspValIleLysAspIleValAspProLysLysLeuGlyTyrThrPheHisGlnSer	487
QY	1816	GAAACAGATAGTAAAGTGTTTGATGCAATCAATCAATGAATATGCAATATGAAGTGAATGA	1877
Db	488	--ThrPheLysAspGlyAlaLysTyrLeuArgGlnLysThrTyrIleGlnAspIle-----	504
QY	1876	AAGATTAAATTCCTCTGCTGCTGATACCGAATAACGCGCTTGATACAAAGCGTCGATTTT	1935
Db	504	-----	504
QY	1936	GAAACCTTTGACCTGAAACAATCTTTTGACGCGGCCATTATTTAAACAATATAAAGTAC	1999
Db	505	-----TyrMetLysGln-----ValAspGlyLysLeuValGlnGlySerProAspAsp	520
QY	1996	TATATATGACGAAGGTACAAATATCGATATATCCGCTTAAACGTAATTTGATGGTGTAT	2055
Db	521	TyrThrAsp-----IleLysPheSerValLysCysAlaGly 532	
QY	2056	GTAATATGATGAATATTTTACTGATGAACTTAATATGAACGTAATTTATTTAAAGAC	2111
Db	533	MetThrAspLys-----IleLysLysGlnValThrPheGlnAsn 545	
QY	2116	GCTAGAGAAATTTTC 2130	
Db	546	PheLysValGlyPhe 550	
RESULT 2			
ID	DPOL.BPM2	STANDARD:	PRT: 572 AA.
AC	P19894:		
DT	01-FEB-1991 (Rel. 17, Created)		
DT	01-FEB-1991 (Rel. 17, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	DNA polymerase (EC 2.7.7.7).		
GN	G.		
OS	Bacteriophage M2.		
OC	Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;		
OC	Phi-29-like viruses.		
OX	NCBI_TaxID=10751;		

```

RN SEQUENCE FROM N.A.
RP MEDLINE=90128268; PubMed=2515115;
RA Matsunoto K., Takano H., Kim C.I., Hirokawa H.;
RT "Primary structure of bacteriophage M2 DNA polymerase: conserved
RT segments within protein-priming DNA polymerases and DNA polymerase I
RT of Escherichia coli."
RL Gene 84:247-255(1989).
CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMATIC ACTIVITIES:
CC DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOLYTIC ACTIVITY THAT
CC Degrades SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -1- MISCELLANEOUS: THIS ENZYME CATALYZES THE FORMATION OF A PRIMER
CC PROTEIN-5'DAMP INITIATION COMPLEX, FOR DNA REPLICATION. IT
CC REQUIRES A PROTEIN AS A PRIMER.
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; J03144; AAA32368.1; -.
CC DR PIR; M00161; JQ0161.
CC DR InterPro; IPR002064; DNA_POL_B.
CC DR InterPro; IPR004868; DNA_POL_B.
CC Pfam; PF03175; DNA_POL_B_2; 1.
CC PRINTS; PR00106; DNAPOLB.
CC SMART; SM00486; POLB; 1.
CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.
CC Transfaser; SMO0116; DNA-POLYMERASE; DNA replication;
CC DNA-binding; Hydrolyase; Exonuclease.
CC VARIANT 192 192 L->F (IN MUTANT APH(2)).
CC FT SEQUENCE 572 AA; 66423 MW; 15E9F27C1632CAC9 CRC64;
CC -----
Alignment Scores:
Pred. No.: 4.66e-07 Length: 572
Score: 234.50 Matches: 139
Percent Similarity: 35.33% Conservative: 97
Best Local Similarity: 20.81% Mismatches: 221
Query Match: 5.76% Indels: 211
DB: 1 Gaps: 32
CC -----
US-09-727-892A-2 (1-2286) x DPOI-BPM2 (1-572)
QY 382 GAGCAGCAGCTTAAGTAAGAGCGCTACTATTGTCGCCAAATCAAAATGTAATT--- 438
DB 11 G|U|T|T|h|T|h|L|y|S|e|u|s|p|r|a|s|p|c|y|a|g|v|a|l|t|r|p|a|l|y|t|g|l|y|t|m|e|t|g|u|l|l|e|g|y 30
QY 439 -----TTAGAAAACGGTTAAATCTTCATCAAT 468
DB 31 A|s|n|l|e|u|s|r|a|s|t|u|l|y|s|l|l|e|g|y|s|n|s|e|r|l|e|u|s|p|l|p|h|e|m|e|t|l|n|t|r|p|a|l|m|e|t|g|u 50
QY 469 TTAGATTTAACATGTTTAAATGGTTTAAATTAAT-----ATTATGTACTATT 522
DB 51 l|e|g|l|a|l|a|s|p|l|e|u|t|y|r|h|e|n|s|a|n|l|e|u|s|p|h|e|a|s|p|g|l|a|l|a|h|e|l|e|a|l|a|n|t|r 70
QY 523 ATGAAAAAC-----ATAACATCA 540
DB 71 l|e|u|g|l|u|g|l|n|s|l|g|y|h|e|l|y|s|t|r|e|a|n|g|l|u|g|l|y|l|e|u|r|o|s|a|n|t|h|y|r|a|e|n|t|r|l|e 90
QY 541 ATTGCACACTTAGTAAAG---AAATTACTGTAT-----GGTGGTATTTTAACAAGATCA 591
DB 91 l|l|e|s|e|r|y|m|e|t|g|l|n|t|r|p|r|y|m|e|t|l|e|a|s|p|l|e|c|y|s|r|h|e|l|y|t|y|r|g|l|y|s|t|a|r|g 110
QY 592 CAACTTAATAACA----- 603
DB 111 L|y|s|l|e|n|h|s|t|h|V|a|l|l|e|t|y|r|a|s|p|r|e|l|e|u|s|l|y|l|e|u|r|p|r|h|e|r|o|v|a|l|l|y|s|l|e 130

```


Db 536 LysGluValThrPheGluAspPheLysValGlyPhe 547

RESULT 4

DPOL_BPCP1 STANDARD; PRT; 568 AA.

AC Q37969;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA polymerase (EC 2.7.7.7).

GN 5

OS Bacteriophage Cp-1.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;

OC Phi-29-like viruses.

OX NCBI_TaxID=10747;

RP SEQUENCE FROM N.A.

RP MEDLINE=95373138; PubMed=7645213;

RA Martin A.C., Lopez R., Garcia P.;

RT "Nucleotide sequence and transcription of the left early region of Streptococcus pneumoniae bacteriophage Cp-1 coding for the terminal protein and the DNA polymerase.";

RL Virology 211:21-32(1995).

RP SEQUENCE FROM N.A.

RP MEDLINE=96211500; PubMed=8648702;

RA Martin A.C., Lopez R., Garcia P.;

RT "Analysis of the complete nucleotide sequence and functional organization of the genome of Streptococcus pneumoniae bacteriophage Cp-1.";

RL J. Virol. 70:3678-3687(1996).

CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMAIC ACTIVITIES: DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOLYTIC ACTIVITY THAT DEGRADERS SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.

CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate + [DNA](N).

CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.

CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch)

CC -----

DR EMBL: 247794; CA87725.1; -

DR InterPro: IPR002064; DNA_pol_B.

DR InterPro: IPR004868; DNA_pol_B_2.

DR Pfam: PF03175; DNA_pol_B_2; 1.

DR PRINTS: PR00106; DNAPOLB.

DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.

KW Transferase; DNA-directed DNA polymerase; DNA replication;

KW DNA-binding; Hydrolase; Exonuclease.

SW SEQUENCE 568 AA; 66148 MW; S15F5EAERF1897896 CRC64;

Alignment Scores:

	Score:	1.59e-06	Length:	568
Percent Similarity:	225.50		Matches:	151
Best Local Similarity:	36.35%		Conservative:	86
Query Match:	23.16%		Mismatches:	258
	5.54%		Indels:	157
DB:	1		Gaps:	32

US-09-727-892A-2 (1-2286) x DPOL_BPCP1 (1-568)

QY 175 GAAGTATTTCCGAGTTTGAATCTTTTATGACGATTTTATACGTATGAAAGACGT 234

Db 38 LysValAsnThrSerLeuGluAspPheLeuLysSerLeuLysLeuAspLeuAspLysThr 57

QY 235 GATACATGCACAAATCAAAACACATATTTATCATGATTTGACACATAACTGTAATAATAC 294

Db 58 TyrThrGluThr---GlyGluAspGluPheLeuLeuPhePheHisAsnLeu---LysPhe 75

QY 295 GATATATCATTTTTCCTTAAGACACCATGGCTATTTTGTATATATACCGGAAAT 354

Db 76 AspGlySerPheLeuLeuSer-----PhePheLeuAsnAspIleGluCys 91

QY 355 ATATATTTAAATCTGCAGAAAGAAATGACACATTTAAATGAAGAGCTACTACT 414

Db 92 ThrTyPhe-----

QY 415 TTAGCCAAAATCAAAATGTAATTTAGAAAACGTGTTAAATCTTCATCAATTTAGAT 474

Db 95 -----IleAsnAspMetGlyValIlePtyrSerIleThrLeuGlu 107

QY 475 TTAACATGTTTAAATGTTTAAATTTATATTTATTTATGATTTATGAACCAAT 534

Db 108 PheProAspPheThrLeuThrPheArg-----AspSerLeuLysIleLeuAsn 123

QY 535 ACATCAATGTCACACATTAGT-----AAGAAATTAAGTGGTGGTATTATTACA 585

Db 124 PheSerIleAlaThrMetAlaGlyLeuPheLysMetProIleAlaLysGly-----Thr 141

QY 586 GAATCACACCTTAAACAGATTTTAATATACGATTTTGGTAAAGTAATGATGAT 645

Db 142 ThrProLeuLeuLys-----His 147

QY 646 GATACGACAGCCTATGACTATGCTGTGAAGTTTGGCAAAACACACCTGAACAACT 705

Db 148 LysProGluVal-----IleLysProGluTyrPhe 157

QY 706 ACATCAATCATTAATGACGTATATATTA-----GGTAGTCCTATTCATTAATAGT 759

Db 158 AspTyrIleHisValAspValAlaIleLeuAlaArgGlyIlePheAlaMetTyrTyrGlu 177

QY 760 GATATATTTCCAAATTTGACTATACAAATTAACATTTTCATTAATTAATGAACT 819

Db 178 GluAsnPheThrLys-----TyrThrSerAlaSerGluAlaLeuThrGluPheLysArg 195

QY 820 TACTTGAATATGAATGACACGTTTTCAGTTTACTCAACCAATATGCAAGATTAATA 879

Db 196 IlePheArgLysSerLysArgLysPheArgPhePheProIleLeuAspGluLysVal 215

QY 880 TCTTATACACTATTCATTCATTCATGATGATGATTTTATGCTATTAATCATCTCT 939

Db 216 Asp-----AspPheCysArgLysHisIle 223

QY 940 CGTGTGCTTAAATATGATATACCAACAATAC-----ATAAACAATTAATT 987

Db 224 ValGlyAlaGlyArgLeuProThrLeuLysHisArgGlyArgThrLeuAsnGluIle 243

QY 988 GATGAGCCTGTTTTCATGATGACATTCGATTCGATTCGATTCGATTCGATTCGAT 1047

Db 244 Asp-----IleTyrAspIleAsnSerMetTyrProAlaThrMetLeuGluAsn 259

QY 1048 AAAATTCACATGTTATATCTTTTACGACACACTATTCAGAACCCAGTTATCCCTACT 1107

Db 260 AlaLeuPro-----IleGlyIleProLysArgTyrLysGlyLys-----ProLys 274

QY 1108 TTTTAAAGATGATGATGATTTTTCATTTATATGATGATGATGATGATGATGATGAT 1158

Db 275 GluIleLysGluAspHisTyrTyr---IleTyrHisIleLysAlaAspPheAspLeuLys 293

QY 1159 -----TTTAAAGATGATGATTTTATTAATAATAATCAACGATGATTCCTCAATGAT 1212

Db 294 ArgGlyLysLeuProThrIleGluIleLysLysLysLeuAspAlaLeuArgIleGlyVal 313

QY 1213 GTAAATACATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1269

Db 314 -----ArgThrSerAspTyrValThrThrSerLysAsnGluValIleAsp 328

QY 1270 ATGATTCACACATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1329

Db 329 LeuTyrLeuThrAsnPheAspLeuAspLeuPheLeuLysHisTyrAspAlaThrIleMet 348

CC phi-29-like viruses.
 OX NCBI_TaxID=10778;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98019084; PubMed=9358052;
 RA Pecenkova T., Benes V., Paces J., Vlack C., Paces V.;
 RT Bacteriophage B103: complete DNA sequence of its genome and
 RL relationship to other Bacillus phages.";
 RL Gene 199:157-163(1997).
 CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMAIC ACTIVITIES:
 CC DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT
 CC DEGRADES SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + (DNA)(N).
 CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X99260; CAA67649.1.
 DR InterPro: IPR002064; DNA_POL_B.
 DR InterPro: IPR004868; DNA_POL_B_2.
 DR Pfam: PF03175; DNA_POL_B_2; 1.
 DR PRINTS: PRO0106; DNAPOLB.
 DR SMART: SM00486; POLBc; 1.
 DR PROSITE: PS00116; DNA_POLYMERASE_B_1.
 KM Transferrase; DNA-directed DNA polymerase; DNA replication;
 KM DNA-binding; Hydrolyase; Exonuclease; Early protein.
 SQ SEQUENCE 572 AA; 66504 MW; 8828E62AB67E309F CRC64;
 Alignment Scores:
 Pred. No.: 3,37e-06 Length: 572
 Score: 220.00 Matches: 122
 Percent Similarity: 35.80% Conservative: 91
 Best Local Similarity: 20.50% Mismatches: 192
 Query Match: 5.41% Indels: 190
 DB: 1 Gaps: 28
 US-09-727-892a-2 (1-2286) x DPOL_BP803 (1-572)
 QY 382 GAACACCATTAATAAGAAAGAGCGTACTATTGACCAAAATCAAAATGTAATT--- 438
 DB 11 GGTGTTTThrThrLeuAspAspCysArgValTrpAlaTyrGlyTyrMetGluIleGly 30
 QY 439 -----TTGAAAACGCTGTTAAATCTTCAATCAAT 468
 DB 31 AsnLeuAspAsnTyrTyrIleGlyAsnSerLeuAspGluPheMetGlnTrpValMetGlu 50
 QY 469 TTGAGTTTAAACAATGTTTAAATGTTTAAATTAAT-----ATTATGTAACCTTT 522
 DB 51 IleGlnIleAspLeuTyrTrpHisAsnLeuLysPheAspGlyAlaPheIleValAsnTrp 70
 QY 523 ATGAAA-----ACCATACATCA 540
 DB 71 LeuGlnIleHisGlyPheLysTrpSerAsnGluGlyLeuProAsnThrTyrAsnThrIle 90
 QY 541 ATTGCAACATAGGTAAAG-----AAATTACTGAT-----GGTGATATTAAAGAAATCA 591
 DB 91 IleSerLysMetGlyGlnTrpTyrMetIleAspIleCysPheGlyTyrLysGlyLysArg 110
 QY 592 CAAGTTAAACA----- 603
 DB 111 LysLeuIleHisThrValIleTyrAspSerLeuLysLysLeuProPheProValLysLysIle 130
 QY 604 -----GATTTTAAATTACGATTTTGTATTAAGATATGATATGATATGATGAGCC 657
 DB 131 AlaLysAspPheGlnLeuProLeuLeuLysGlyAspIleAspTyrHisAlaGluArgPro 150

QY 658 TATGACTATGCTGTGAATGTTTGGCAAAACGACACGTAACAACTTACATTCAT 717
 DB 151 ValGlyHis-----GluIleThrProGluGluTyrGluIleLys 164
 QY 718 AATGAGTATATATATATGATATGTCATATTCATTAATAGCATATATTCCAATTT 777
 DB 165 AsnAspIleGluIleIleLeuAlaArgala-----LeuAspIleGlnPheLysGln 180
 QY 778 GACTATAACAATTAACATTTTCATTCATGATATATATGATATCTTACTGATATATGAAG 837
 DB 181 GlyLeuAspArgMetThrIleAspSerLeuLysGlyPheLysAspIleLeuSer 200
 QY 838 ACACGT-----TTTCAGTTACTCAACCAATACAGATATTAATATATCT 882
 DB 201 ThrLysLysPheAsnLysValPheProLysLeuSerLeuProMetAspLysGlu----- 218
 QY 883 TATACATATATCATTTCCATGATATGAATTTTATGACTATATTAATCATCTATCT 942
 DB 219 -----IleArgAlaTyrArg 224
 QY 943 GGTGTTTAAATATGATATACCAACAAATACATTAACAACTATGATGAGCTGTTT 1002
 DB 225 GlyGlyPheThrTyrLeuAsnAspLysTyrLysGluLysGluIleGlyGluLys---Met 243
 QY 1003 TCTATTGACATCATTCGATTCCTTATGTCATGATCATGAAAAAATTCACAACTGG 1062
 DB 244 ValPheAspValAsnSerLeuTyrProSerGlnMetLysSerArgProLeuPro----- 261
 QY 1063 TTATATTTTACGAACACATATCAGAACCAAGTTATGCCACTTTTATGATGATGAC 1122
 DB 262 -----TyrGlyAlaProIleValPheGlnGlyLysTyrGluLysAsp 275
 QY 1123 AATTATTTTTCATATATATAGATATGATATGATATTAACGATGATTA----- 1173
 DB 276 GluGlnTyrProLeuTyr---IleGlnArgIleArgPheGluPheGluLeuLysGluGly 294
 QY 1174 -----TTATTAATTAATTAATCCGCTGATTACGTCAAATGATTGATAAATACAT 1224
 DB 295 TyrIleProThrIleGlnIleLysLysAsnPro-----Phe 306
 QY 1225 AATAATGATATGATATGATATATATATACATACAAATACATTAAGATGATTAAGACAT 1284
 DB 307 PheLysGlyAsnGluTyrLeuLys---AsnSerGlyAlaGluProValGluLeuTyrLeu 325
 QY 1285 ACGGATATGATTCATGATATACGCTGTTAATTCGTTTATATATGATGATGATAC 1344
 DB 326 ThrAsnValAsp---LeuGluLeuIleGlnIleHisTyrGlnMetLysAsnValGluTyr 344
 QY 1345 -----TTTCATGACGCTGATATTAATTTTCAAAACTAT----- 1377
 DB 345 IleAspGlyPheLysPheArgGluTyrThrGlyLeuPheLysGluPheIleAspLysTrp 364
 QY 1378 ---TTTATTAACACACAGTAAGTTAAAAAACAATCAATATGACATCACCTTAACGAC 1434
 DB 365 ThrTyrValLysThrHisLysGlyAlaLysLys----- 376
 QY 1435 TATCATATTAACGATATCAACAGAACCCATATCATCAATAGAGAGGTATGTTATCT 1494
 DB 377 -----GlnLeuAla 379
 QY 1495 AAGTCGTTTAAATGATATATGCG----- 1521
 DB 380 LysLeuMetPheAspSerLeuTyrGlyLysPheAlaSerAsnProAspValThrGlyLys 399
 QY 1522 ATAACCTGATTAACGTCATATTTACTTA---TTCCGTTTATGATGATATACATGACATA 1578
 DB 400 ValProTyrLeuLysGluAspGlySerLeuGlyPheArgValGlyAspGluGlu----- 417
 QY 1579 TACAATATCATTAACGTCACAAAAC-----ACGAGACGTAATATATTA----- 1623
 DB 418 -----TyrLysAspProValTyrThrProMetGlyAlaPheIleThr 431

QY 1624 -----TTCCTCATTTGTGCATGCTTCATTTGATTAATGTTGCTCT 1671
 Db 432 AATATPALAARPhenhrThrlleThralaAlaGlnAlaCysTyr----- 446
 QY 1672 TTCATACTTAACGGAAGTAATGACACATTTATTTATTCGATGATAGT 1731
 Db 447 -----AspArgllelleTyrCysAspThrSper 456
 QY 1732 TTGTATGAAATCCGTTGTAAACCCCTTATTAACCCAGTTATTCGACCCGATAGCC 1791
 Db 457 lIeHlsleuThGlyThGlnuAlProGluIlellelYsAspIleValAspProLysLys 476
 QY 1792 TTAGGAAATGGGATATGAAACGACAGATAGATAGATGTTGTCTGATCATGATAG 1851
 Db 477 LeGlyTyrTThPAlahlsGlsuSer---ThrpheLysArgAlaLysTyrLeuArgGlnLys 495
 QY 1852 AAA-----TATGCATATGAGTGAATGGAAGATT 1881
 Db 496 ThTyrllleGlnAspIleTyrAlaLysGlnuValAspGlyLysLeu 510
 RESULT 7
 ID BXEN_CLOBU STANDARD; PRT; 1162 AA.
 AC Q06366;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Botulinum neurotoxin type E, nontoxic component.
 OS Clostridium butyricum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 ON NCBI_TaxID=1492;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BL6340.
 RX MEDLINE=93360835; PubMed=8355622;
 RA Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
 RA Takeshi K., Ohyama T., Isogai E., Isogai H.;
 RT "Similarity in nucleotide sequence of the gene encoding nontoxic
 RT component of botulinum toxin produced by toxigenic Clostridium
 RT butyricum strain BL6340 and Clostridium botulinum type E strain
 RT Mashike";
 RL Microbiol. Immunol. 37:395-398(1993).
 CC - FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
 CC TOXICITY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D12739; BAA02231.1;
 CC InterPro: IPR000395; Bontoxilysin.
 CC Pfam: PF01742; Peptidase_M27.1.
 CC PRINTS: PR00760; BONTOXILYSIN.
 CC ProDom: PD001963; Bontoxilysin. 1.
 CC Neurotoxin.
 CC SEQUENCE 1162 AA; 136829 MW; C86E9BE932DA78E4 CRC64;
 Alignment Scores:
 Pred. No.: 1.55e-05 Length: 1162
 Score: 208.50 Matches: 196
 Percent Similarity: 34.22% Conservative: 150
 Best Local Similarity: 19.39% Mismatches: 291
 Query Match: 5.12% Indels: 374
 DB: 1 Gaps: 55
 US-09-727-892a-2 (1-2286) x BXEN_CLOBU (1-1162)
 QY 103 AACCACCAATATATAAAACGTACTTATCTGTAGCAATGTTGTTAATGTTAT 162

Db 203 LysProAsnAspAsnLeuAsnIleProTyrArgLeuArg---AsnGluPheAsnSerLeu 221
 QY 163 GAAATGATGTTGAAGTATTTCCGACTTTGAAATCTTTTATGACGATTTATACGAT 222
 Db 222 Glu-----TyrSerGlu 225
 QY 223 GTCAAAAGACGTGTATACATATCAAAATTCAAAACAGATATATCATGATGACATAC 282
 Db 226 LeuAspMetlleAspPheLeuIleSerGlyGlylleAspTyrLysLeuLeu---Asn 243
 QY 283 TGATATTAATAC-----GATATCATTTTACTTAAAGACACACGCTTATTT 333
 Db 244 ThrAsnProTyrTThPpelleAspLysTyrPhele-----AspThrSerLysAsnPhe 261
 QY 334 GATATATATACACCGGAAATATATATTTAAATCTGCAGACAGAAATGACACACATTA 393
 Db 262 GluLysTyr-----LysAsnAspTyrGluIle 270
 QY 394 AAAATGAAAGAGCTACTATTTTACCAAAATGCAAAATGTAATTTTGAAGAAACGTGT 453
 Db 271 LyslleLysAsnAsnAsnTyrIleAlaAsnSerlleLysLeuTyrLeuGlnGlnLysPhe 290
 QY 454 AAATCTTCAATC-----AATTGATTTAACAAGTTTAAATGGTTT--- 498
 Db 291 LyslleAsnValLysAspIleTPrGluLeuAsnLeuSerTyrPheSerLysGluPheGln 310
 QY 499 -----AAATTTAAT 507
 Db 311 lleMetMetProGluArgTyrAsnAsnAlaLeuAsnHsIstYrTyrArgLysGluPheTyr 330
 QY 508 ATTATGATTAATCTT-----ATGAAACCC 531
 Db 331 VallleAspTyrPheLysAsnTyrAsnIleAsnGlyPheLysAsnGlnlleLysThr 350
 QY 532 AATCATCATTCGACACATTAGTAACAATTACTT----- 567
 Db 351 LysLeuProLeuSerLysTyrAsnLysGluIlelleAsnLysProGluLeuIleValAsn 370
 QY 568 -----GATGGT----- 573
 Db 371 LeuIleAsnGlnAsnAsnThrValleuMetLysSerAsnPheTyrGlyAspGlyLeuLys 390
 QY 574 -----GTTATTTACAGAACATCACAACCTTAACAAGAT--- 606
 Db 391 GlyAsnValAspAsnPheTyrSerAsnTyrIlelleProTyrAsnLeuAsnTyrGluHis 410
 QY 607 ---TTATATATAGCATTTTGTATAGATATAT-----GAT 639
 Db 411 SerlleAsnTyrSerTyrLeuAspAsnValAsnIleGluIleGluLyslleProPro 430
 QY 640 ATGATATATAGTGAAGCCTATGACTAT----- 666
 Db 431 lleAsnAspGluAspIleTyrProTyrArgLysAsnAlaAspThrPheIleProValTyr 450
 QY 667 -----GCTGCAATGTTTTCGAAACCTGACACCTGACACATGATCATTCAT 717
 Db 451 AsnIleThrLysAlaLysGluIleAsnThrThrThrProLeuProValAsnTyrLeuGln 470
 QY 718 AATGACGCTGATTAATATGATGATGCGCATATTCATTAATGATATTTCCAAATTTT 777
 Db 471 AlaGlnMet-----lleAspSerAsnAspIleAsnLeuSerSer 483
 QY 778 GACTATTAACAATTAACATTTTCATTTG---AATATATGCAATCTACTGATATATGAA 834
 Db 484 AspPheLeuLysValIleSerSerLysGlySerLeuValTyrSerPheLeuAsnAsnThr 503
 QY 835 ATGACAGCTTTTCAGTACTACCAACCAATATCAAGATATTAATATCTTATACACATTAAT 894
 Db 504 MetAspTyrLeuGlnuPheIlelleLysTyrAspLysProIleAspThr----- 518
 QY 895 CATTTCCATGATATGAATTTTATGACTATATTAATCATCTATGCTGCT----- 945

RC STRAIN-ATCC 24666 / AG4;
 RX MEDLINE=91347410; PubMed=1879001;
 RA Robinson M.M., Royer J.C., Horgen P.A.;
 RT "Homology between mitochondrial DNA of Agaricus bisporus and an
 internal portion of a linear mitochondrial plasmid of Agaricus
 bisporus".
 RL Curr. Genet. 19:495-502(1991).
 CC -1 CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + (DNA)(N).
 CC -1 MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
 CC (BY SIMILARITY).
 CC -1 SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG TO
 CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: X65075; CAA44800.1; -
 CC PIR: S28103; S28103.
 CC InterPro: IPR002064; DNA_pol_B.
 CC InterPro: IPR004868; DNA_pol_B_2.
 CC Pfam: PF03175; DNA_pol_B_2; 1.
 CC SMART: SM00486; POLBc; 1.
 CC PROSITE: PS00116; DNA_POLYMERASE_B; FALSE_NEG.
 CC Transfaser: DNA-directed DNA polymerase; DNA replication;
 CC DNA-binding; Plasmid; Mitochondrion.
 CC FT NON_TER 1
 CC SQ SEQUENCE 797 AA; 91922 MW; 5C49EAF51FB7927 CRC64;
 Alignment Scores:
 Pred. No.: 1,69e-05 Length: 797
 Score: 208.00 Matches: 165
 Percent Similarity: 37.18% Conservative: 125
 Best Local Similarity: 21.15% Mismatches: 248
 Query Match: 5.11% Indels: 242
 DB: 1 Gaps: 45
 US-09-727-892a-2 (1-2286) x DPOM_AGABT (1-797)
 QY 31 AAGCATGAGCTGATGATTTTATCTGATATGAGAACATGACGTACATAAAGT 90
 DB 146 LYSHISASPLEUARG---ILEGLYSERLEUASPLEUGLUTHR----- 158
 QY 91 AAGGAGCGAAAAAACCACCAATATATAAACGTTACTTATTCGTACCAATGCTTGC 150
 DB 159 -----TyrGlyAspAsnCysPheGluLeuGlyLeuGlyAsn 170
 QY 151 TTATAGTGTATGAAATGATGATGATTCGAGATTTCGAAATCTTTAT----- 204
 DB 171 LeuAsnValTyrAlaGlyGlyPheAlaLeuAsnAspGlyPheLysLysLeuTyrLeu 190
 QY 205 -----GACGCA 210
 DB 191 AsnAsnAspThrGluLeuAsnSerGlyAlaIleLeuLysMetPheGlyLeu 210
 QY 211 TTTATACGTATGTAAGAGCGTATACATCAACAATAAAGATATTTATCAG 270
 DB 211 PheAspTyrIleAlaGluAspArg-----LysAlaArgAsnSerGlyThrIle 226
 QY 271 ATTGCACATAGCTGTAATATAGATATATCTTTTACTTAAAGACCATG----- 324
 DB 227 TTTAAHISASPLEUGLYARGPheAspSerValPheIleIleArgSerLeuGlySerGlu 246
 QY 325 -----CGTTATTTGATATATATATATACACGGGAATAATATATTTAAA 366
 DB 247 GlyTyrLysIleAsnGlyLeuIlePheAspAsn-----SerIleLeuTyrLeuLys 263
 QY 367 TCTGCAGAGAAATGACAC---ACATTTAAATGAAAGAGGCTACTATTTTACC 423

DB 264 IleValAspSerThrArgLysLeuThrIleLysLeuAspSerIleLysLeuValPro 283
 QY 424 AATCAAAAATGATATTTTACAAAAGCTGTTAAATCTTCA---ATCAATTTACATTTAACA 480
 DB 284 HisSer-----LeuAspLysAlaLeuSerSerAsnGlyCysAsnIleSerLysGly 300
 QY 481 ATGTTTAAATGATTTTAAATTAATATATATGATTAATGATAAACAATACATCA 540
 DB 301 MetPheProHisLysPhe----- 306
 QY 541 ATTGCACATTTGGTAAATTAATCTGATGCTGTTATTTACAGCAATCAACTTAA 600
 DB 307 -----ValAsnLysAspThrLeuAsn-----TyrIleGlyAsp-----Lys 318
 QY 601 ACAGATTTTAATTAATACGTTTGTATTAAGATATATGATATGATGATGAGGCTAT 660
 DB 319 ProAspIleLysTyrTyrVal-----AspGluAsnLysPheAsnGlySerLysLys 336
 QY 661 GACATAT---GCTGTGAAATGTTTTCGAAAACTCACACCTGAAACAATTCATATTCAT 717
 DB 337 LysTyrLysSerLeuProSerIleLeuAsnLeuLysGlyLysGlyAsnTyrLeuAsp 356
 QY 718 AATGACGTATATATTTAGTATGTCATATTCATTTATGATATATTTCCAAATTTT 777
 DB 357 LysAsp-----IleLeuGlyLeuLeuGluLeu----- 365
 QY 778 GACATTAACAATTAATACATTTTCATGAAATATTAAGAACTTACTTGAATATGAAATG 837
 DB 366 ---MetAsnLysVal-----SerLeuThrLysPheAsnGlyTyrLysLeuAsn---Ile 381
 QY 838 ACACGTTTTCAGTACTCTACACCAATATCAAGATATTTAAATATCTTTATACATATTCAT 897
 DB 382 ThrLysPheSerThrLeu-----ProSerIleThrLeuAsnIlePheGlyIleArg 398
 QY 898 TYCATATGAT-----ATGAAT-----TTTATGACTATATTTAA 930
 DB 399 PheTyrAspAspGlnAsnSerIleLysMetIleAsnGlyProLeuSerGluPheIleArg 418
 QY 931 TCATTTCTTCGTCGTGCTTAAATATGATTAACAACCAATACATTAACATTAATGAT 990
 DB 419 SerSerTyrPheGlyGlyAsnSerAspIlePheValSerGlyGluArgLeuValLys 438
 QY 991 GAGCCTGTTTTCATGATCATCATCATGAGTATTCCTTTGATGATATCATGAAATA 1050
 DB 439 AsnGly---TyrHisTyrAspMetAsnSerGlnTyrProTyrAlaMetLeu---GlnSer 456
 QY 1051 ATTCACACATGG-----TTATACCTTTTACGACAC--- 1080
 DB 457 MetProThrGlyAsnProValPheSerThrAsnThrAspLeuAsnTyrTyrArgAsnGly 476
 QY 1081 -----TATTCAGAACCAAGCTTAATCCCTACTTTTAA--- 1113
 DB 477 PheValPheAlaArgValThrProProSerLysAspThrLeuValAsnLeuPheIlePro 496
 QY 1114 -----GATGATGACATATTTTTCATATATATAGATGATTAAGATGATATTTAAC 1164
 DB 497 ArgArgSerAspArgGlySerValIle-----CysAspArgAsnThrPheTyr 512
 QY 1165 GATGATTTATTAATTAATTAATTAATCAACGATATAGCTCAATGATGTAATAATCTAT 1224
 DB 513 Glu-----PheIleProThrProAspLeuLysGlnGlyLeuGlyTyr 528
 QY 1225 AATAATGAT----- 1233
 DB 529 LysPheGluValIleCysGlyIleAsnProPheProAspValLysGlyAsnGlyLeuPhe 548
 QY 1234 -----AATGATTCAGTTTAATTAATTAATCAATAATCAATTAAGATGATTCAGAC 1281
 DB 549 SerGluPheValAsnHisPheTyrGluIleLysSerSerThrAspLeuGlyGlnLys 568
 QY 1282 ATTCAGGATATGATTCAGATGACATATACGCTGTTGTTATATGATGATGAA 1341

```

Db 569 TyrIleAlaLysLeuSerLeuSerLeuTyrGlyLysPheGlyGlnLysGluArgGlu 588
QY 1342 TACTTT-----CATGCACCGCATATATTTTTCACAAACATATTTT 1380
Db 589 TyrSerIleArgLeuLeuGlnLysAspLysAlaLysGlnIleIleSerLysAsnHisTyr 608
QY 1381 -----ATTAAACACACAGTAAGTTTAAA 1404
Db 609 SerTyrMetSerGlnValSerAspAsnTyrThrLeuIleLysSerGlyGlyArgLeuAsn 628
QY 1405 AACAAATCATATGACATCATCCTTACGACTATACATATGATATGACACGACAC 1464
Db 629 SerLysLeuArgArgLeu-----TyrAlaGlnGlnAlaArgIleAsnThr 643
QY 1465 CCATCTCAATGACGAGGTATGATATCTAAGTCGTTTAAATGATATATGCGATA 1524
Db 644 -----IleAsnAspSerLeuLeuSerSerLysPheIleLysSer-----ArgGlyIle 659
QY 1525 CCGTCATTTACGTTACATTTTAACTATTCGTTTGAATGATACAAATGACATATCAAT 1584
Db 660 ProSerAla-----ValGlnIleSerAla 667
QY 1585 ATCATTAACGCTTACAAAACACTGACGTAATATATATATCTCTCATATTTGTCATCA 1644
Db 668 MetIleSerSerTyrAlaArgThrSerIleAsn----- 678
QY 1645 CGTTTCATTTGATTAACCTATTTGTTCTTCCATTTACTTACGGAAGTAAGTAAGTACGAC 1704
Db 679 -----ProPheLys-----AsnIleProGly 685
QY 1705 AATTTT---ATTATTCGCTACTGATGATTTGTATATGAAATCCGTTTAACCTTA 1761
Db 686 AsnLeuAlaIleAlaSerAsnThrAspSerLeuIleLeuArg-----LysProLeu 702
QY 1762 TTGACCCGAGTTTATTCGACCCGATAGCCTTAGGTAATGAGTAATGAAACGACAG 1821
Db 703 GlnAspHisLeuIleGlyLysGlnIle-----GlyLysThrPylsLeuGlnHisLysPhe 720
QY 1822 ATAGATTAAGATGTTTACTGATGATCATAGAAATATGACATATGA-----GTGAT 1872
Db 721 LysAsnGlyValPheVal---LysProLysLeuTyrCysTyrGlnAspValAspIleAsn 739
QY 1873 GGAAGATTAATTCGCTTCGCGATACCGGAAACCCCTTTGATACAGCGTGCAT 1932
Db 740 GlnLeuIleArgLysAlaSerGlyValThrAlaSerAsn-----LeuThr 754
QY 1933 TTGGAACCTTTTACGCTGGAACAATCTTTGACGCTCCATATTTGAAACATATAAGT 1992
Db 755 TyrGlnAsnPheVal-----GlnLeuValAsnGlyLysAspValIleuThrAsnLysGln 772
QY 1993 ATCTATTAAT-----GAGCAAGGTACAAATATGCGATATATCCGCTTAACAGTAAT 2043
Db 773 LeuPheArgLeuAsnThrPgluThrLeuAsnIleGlnIleValAsnIleAsnThrLysIle 792

RESULT 9
YCF2_MARPO STANDARD; PRT; 2136 AA.
ID YCF2_MARPO
AC P09375.
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 259 kDa protein ycf2 (ORF 2136).
GN ycf2.
OS Marchantia polymorpha (Liverwort).
OC Chloroplast
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiales; Marchantiaceae; Marchantia;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyanagi K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,

```

```

RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT Liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC - SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X04465; CAA28078.1; -.
CC PIR: A05037; A05037.
CC PIR: S01591; S01591.
CC InterPro: IPR003959; AAA_Atpase_cent.
CC Pfam: PF00004; AAA; 1.
CC Chloroplast; Hypothetical protein.
KW SEQUENCE 2136 AA; 259911 MW; 5BD170C9C6F61197 CRC64;

Alignment Scores:
Pred. No.: 1,71e-05 Length: 2136
Score: 207.50 Matches: 189
Percent Similarity: 35.88% Conservative: 163
Best Local Similarity: 19.27% Mismatches: 293
Query Match: 5.10% Indels: 336
DB: Gaps: 54

US-09-727-892a-2 (1-2286) x YCF2_MARPO (1-2136)
QY 64 ATGGAACATTTAGCGTACATATAAGTTTAACGACGAAAAACCAACCAATATAAAGC 123
Db 324 IleGlnSerPhePheLeuLeuLysIleLysGlyAsn-----LeuTyrPheLysAsn 340
QY 124 GTTACTTATTCGTACCAATGCTTGG----- 150
Db 341 -----TyrIleGlnPheValThrTrpGlnSerTyrLysLysAspCysLeuAspPheAsn 358
QY 151 -----TTATAGCTTATGAATATGATGTTGAAGTATTCGAGTTTCGATCT 198
Db 359 LysPheAsnGlnLeuLeuAsnAsnSerGlnIleTyrIleLysIleGlnGlnLeuPheAsp 378
QY 199 TTTTATGACCATTTTATACGTATGTAAGACGATACAAATCACAATCAAAATCAAAACA 258
Db 379 TyrIleTyrLysPheSerLysTyrIleLeuTyrGlnGly-----LysLysSerLysThr 396
QY 259 GATATTAATCATGATTCACATACATCTATTAATACGAT-----AATCATTTTTRA 309
Db 397 ---IleIleLysIleSerPheAsnAsnAsnIleTyrTyrLysLysLeuAsnSerIlePhe 415
QY 310 CTTAAAGACACCATCGCTTATTTGATTAATATAC-----CGCGAA 351
Db 416 AsnPheAsnThrIlePheTyrPheAspSerAsnAsnLeuLeuPheAspThrLeuLysIle 435
QY 352 AATATATATTAAAA-----TCTCGAAGAAGA 378
Db 436 AsnTyrTyrIleAsnAsnLysProPheLeuLysSerPheLeuIleTyrSerIleSer 455
QY 379 AATGAACACATTTAAATGAAGAGCGTACTATTTTACCCAAATGCAATGTAAT 438
Db 476 -----LysLysAsnSerLysAspValIleThrAsnValPheSerLysGlnAsn 491
QY 439 TTGAAAAAACGCTTAATCTTCAATCATTTGATTTAAACATGTTT----- 486
Db 476 -----LysLysAsnSerLysAspValIleThrAsnValPheSerLysGlnAsn 491
QY 487 -----TTAAATGGTTTT-----AAA 501
Db 492 LysIleGlnIleAsnAsnPheSerLysSerIleTyrTyrAlaPhePheGlnIleLeuSer 511
QY 502 TTTAATATTTATTTGATAC---TTATGAAACCAATCATCAATTCGACATTTAGTAAG 558

```

Db 512 ILeaSnGluLeaSpAsnLysPheValIleAsnLysIleSerLeuLysAsnIleAsnLys 531
QY 559 AAATACTGATGCGTGGTTATTTAACAGAAATCAACAATTAAGATTTTAAATATACG 618
Db 532 LysLysGlnLysArGpHeYrLeuAsnLysIleLysSerSerAspSnpHeArGpHe 550
QY 619 ATTTTGATTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 678
Db 551 -----IleAsnLeuTyrLysIleLysAsnTyrIleSerGln 563
QY 679 TTTCGAAA-----CTCAGACCT-----GACCACTTACA 708
Db 564 PheValSerAsnSerPheLeuLeuAsnProAlaPheGluLeuGlnGlnAsnTyr 583
QY 709 TACATTCATATACGCTGATATATATAGTATGCTATTCATATATATGCTATATTT 768
Db 584 TyrLeuLysLysLysAsnIleLeuPhe-----PheLysLysLeuAsnGlnValPhe 600
QY 769 CCAAAATTTGACTAT----- 783
Db 601 SerAsnPheHeYrPheGlnTyrTyrLysCysLysLysLeuAsnIlePheLeuLysPhe 620
QY 784 -----ACAATTAACATTTTCATGATATTT 810
Db 621 AlaSerLeuGluLysIleLeuLysLysArGAsnLysLysPheThrIleSerIleLysLeu 640
QY 811 ATGGAATCTTACTGATATATGATGATGATGATGATGATGATGATGATGATGATGATG 849
Db 641 PheLysLysPheYrLysAsnLysLeuAsnGlnLysGlnTyrLysIleLysLeuSerGln 660
QY 850 TTATCTACCAATATCAACAATATTT-----AAATATCTTATACCAATATATCATGATG 906
Db 661 IleLeuGlnAsnGlnLysLysLysLeuAsnLysLysArGAsnLysLysPheGlnPheAsnPro 680
QY 907 -----ATGAATTTTATGAC-----TATATTAATCATTC 936
Db 681 AsnIleLysIleLeuSerPheYrAsnSerSerLysLysAsnIleTyrLeuGlnAsnLys 700
QY 937 TATCGGTGGTGTAAATATGTTAAACACCAATATACATA-----AAC 978
Db 701 TyrPhePheAsnLysAsnLeuIleAsnLysLysIleThrTyrLysLysIleSerAsn 720
QY 979 AAACATATGATGAGCCTGTTTCTATGATGATGATGATGATGATGATGATGATGATGATG 1038
Db 721 LysLeuValIleSer-----AsnSerGluTyrAsnLysIle 733
QY 1039 TATCATGAAAAAATCCACATGCTTAACTTTTACGAACACTATTCAGAACCAACGTTA 1098
Db 734 TysAsnLysLysAsn-----MetLysPhePheSer-----PheSerLysAsnSerVal 749
QY 1099 ATCCCTACTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1138
Db 750 LeuAspThrPhePhePheAsnLysLysSerPheAsnIleIleThrVal-----Ile 766
QY 1159 TTTAAGATGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1212
Db 767 PheAspLysLeuLysLysIleGlnLeuAsnPheGlnGluIleGlnLysIleLeuAsnGly 786
QY 1213 -----GTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1239
Db 787 PheSerLeuPhePheAsnSerLysAsnIleLysLysThrLysIlePheLysAsnSerTyr 806
QY 1240 TACGTTAATATCAATACAAATACATTAAGATGATGATGATGATGATGATGATGATGATG 1287
Db 807 PheIleAsnGlnLysLeuThrThrThrPheSerPheAsnAspLysGluPheAsnIlePhe 826
QY 1288 GGTATGATTCGATGCAATATACGCTTATATTCG----- 1320
Db 827 PheLeuGluLeuPheIleSerGluIleAsnAspPheLeuMetArGpPheLysLys 846
QY 1321 TTTCGTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATTTT 1368
Db 1321 ----- 1368

Db 847 TyrLeuTyrTyr-----ArgIleTyrLysAspLysGluIleLeuPheAsnProIleGlu 864
QY 1369 -----CAAACTATTTTATTAATAACACAGGTAAGTTAAAAACAAATC 1413
Db 865 AsnArGlnLeuLeuGlnAsnPheGlnLysThrLys----- 877
QY 1414 AATATGATCATCGCTTACGATCATCATCTACTGATGATATCAACGAACCACTACTCA 1473
Db 878 -----IleLeuThrPheIleAspPheLeuGlnAspProGluLeuAsn-----TyrAsn 893
QY 1474 AATGAGAGGTTATGTTATCAAAAGCTTTAAATGATATATGATATGATATGATATGAT 1533
Db 894 AsnArGpHeIlePhe----- 898
QY 1534 CGTTCACTTTTAACTATTTCCGTTTATGATGATATCAACATGATATC-----AATATCAT 1590
Db 899 -----HisLeuGluLysLysThrIleLysAsnAsnLeuLeuTyrLeuArGLeuVal 916
QY 1591 AACGTTCAAAAACACTGACGTAATATATTA-----TTCTGATCTTGTGTC 1638
Db 917 LysIlePheLeuLysAspLysArGAsnPheLeuLeuIleAsnGluIleLysSerPheIle 936
QY 1639 ACATCAGCTGATGATATTAATATATGTTGCTTCCATATCTTAACGAAAGTGAATTT 1698
Db 937 GluLysLysAsn-----AsnLeuPheIleLysSerGlnLeuSerAsnValLeuVal 954
QY 1699 GACGACATTTTATTTATGCTGATCT-----GATAGTTGTTATATGAATTCGTTGTT 1752
Db 955 LysAsnSerTyrLysPhePheAspAsnIlePheAsnPheHisPheLysGlnLysGlu 974
QY 1753 AAACCC-----TATTAACCCAGTTTATTCGACCCGATAGCTTATAGTAA 1800
Db 975 LysAsnIleGluIleIleLeuAsnAsnGlnAsnTyrPheGluLysSerLeuLeuLys 994
QY 1801 TGGGATATGAAACGAACAGATAGAT-----AAGATGTT 1836
Db 995 ThrTyrLeuLysAsnLeuAsnLeuAsnSerTyrSerLysPheSerTyrLysIlePhe 1014
QY 1837 GTA----- 1839
Db 1015 IlePheGlnLeuLeuAsnIleLeuAsnLysAsnAsnTyrLysThrPheGlnThrPheIleSer 1034
QY 1840 -----CTGATCATATAG----- 1851
Db 1035 GluLeuIlePheYrSerLysAsnLeuAsnTyrLysIleGlnAsnLysIleGluLysAsn 1054
QY 1852 AAATATGCATATGAAATGAT-----GGAAGATTAATTAATGCTGCTGCT 1896
Db 1055 AsnTyrCysTyrAsnLysAsnIleSerTyrLysLysLysLysIleYrThrValAsnPhe 1074
QY 1897 GGTATACCGAAAAAGCGCTTTGATACAGCGTCGATTT----- 1935
Db 1075 PheGluLysAsnLeuPheGlnThrAsnAsnSerTyrPhePheThrLeuGluTyrTip 1094
QY 1936 -----GAAACCTT----- 1944
Db 1095 GluTyrAsnThrTyrIleLeuLeuGlnIleIleGlnGluThrPheGlnIleThrAsp 1114
QY 1945 GTACGTGCAACATTTCTTGAAGGTGCGCATTTGAAACAAAT----- 1986
Db 1115 ValLeuGluTyrPheLysLysLysIleIleGluLysAsnLeuLysPhePheLeuLys 1134
QY 1987 -----AAAGTATCTATATATGCAAGGACAAATATCGATATATCCGCTAAACTGAATTT 2043
Db 1135 SerLysLysIle-SerLeuLysThrLeuSerPheHisAsnPheLysLeuLysTyrAsnLe 1154
QY 2044 G-----TATGTGTAATGATATGATGATGATGATGATGATGATGATGATGATGATG 2077
Db 1154 uArGpPheAsnGlnLysAsnTyrLysLysAsnTyrLeuLeuAsnPheLeuTyrPheSerAs 1174
QY 2078 -ATGACTTATATGAAACGTC--AATTTATATTAAGACGCTAGACAAAATTTGAC 2133
Db 1174 pPheAsnLeuIleAsnAsnGlyAsnAsnLeuTyr-----TyrVal 1187


```

OY 2134 CATGTCGATTTGGATTCATTTCTTAATATG-----AAAGCAATCGGTT----- 2179
Db 1187 lilepseriesleuValllephelielepneleuTYTTCylnlystlelepheserlellell 1207
OY 2180 -----CATTTGCACCTTAAGCACTTAATTCGACGTGAACGTTGACATACAA 2229
Db 1207 eglySeraspScyspheHisleuTrpLysAsnphelguile---lleglnTYTleuThrAs 1226
OY 2230 T 2230
Db 1226 p 1226

RESULT 10
RPOB_PLAFA
ID RPOB_PLAFA STANDARD; PRT; 1024 AA.
AC P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BW(C10);
RX MEDLINE=95107345; PubMed=7808472;
RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RA Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RT "Phylogenetic analysis of the rpoB gene from the plasmid-like DNA of
RL Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE=91187055; PubMed=2011147;
RA Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT "A circular DNA in malaria parasites encodes an RNA polymerase like
RL that of prokaryotes and chloroplasts."
RL Mol. Biochem. Parasitol. 44:115-124(1991).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -1- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL, X75544; CAA53332.1; -
DR EMBL, X52177; CAA36427.1; -
DR EMBL, X95275; CAA64572.1; -
DR PIR, S10438; RNZOBF.
DR InterPro: IPR001572; RNA_POL_B.
DR Pfam: PF00562; RNA_POL_B; 1.
DR PROSITE, PS01166; RNA_POL_BETA_1
DR Transferrase; Transcription; DNA-directed RNA polymerase.
KW SEQUENCE 1024 AA; 122185 MW; B49FBDEB7C9BD412 CRC64;
SQ

Alignment Scores:
Pred. No.: 1.91e-05 Length: 1024
Score: 207.00 Matches: 170
Percent Similarity: 41.52% Conservative: 146
Best Local Similarity: 22.34% Mismatches: 232
Query Match: 5.09% Indels: 215

```


109 ThrAlaSerAsnPheserGlnAsnSerLeuProSerLysIleSerLysAsnIlePhePhe 128
433 -----GTAATTTAGAAAA----- 447
129 IleuStyrProIleIleLeuLysIleHisAlaAspLysIleLeuPheThrSerPro 148
448 CGCTTAATCTTCATCAATTTAGATTTA----- 477
149 LysValArgValSerPhe-----LeuAspValLeuSerLysIleLysLeuValArgAsnAsn 167
478 -----ACAAATGTTTAAATGGTTTAAATTTAT--- 507
168 IleIlePheSerProThrTyrIleAsnThrIleHisValSerSerAlaLysPheAsnPhe 187
508 -----ATTATGTACTACTTTATGAAGAACCAATACA 537
188 GlnLysLysAsnIleLeuAsnLysSerThrIleIleLysAsnPheAsnLysIle----- 205
538 TCATTTGCAACATTAGGTAGAATTAATCTTGATGGTATTATTAACAGATCAACAACT 597
206 -----LysLysIleLysSerPheSerLysIlePheSerSerAsnGlnThr 219
598 AAAACAGATTTTAATTTATACGATTTTGTATTAAGATATATGATATGATAGTGAAGCC 657
220 LysLysAsnPheProLeuAsnIleTyrLeuLysSerLeuLysCysAsnLysThrGlnPhe 239
658 TATGCTATGCTGTGAATGTTTGGCAAACTCACACCTGCAACAACTTACATCATTCAT 717
240 IleAspTyrGlnTyrLysAsnLeu-----LeuGlnValGlnLeuGlnAlaAsnIleGln 257
718 AATGACGTATTATTAGTATGTATGTCATATTCATTATAGTATATATTTTCCAAATTTT 777
258 AsnAsnIleLeuGlnIleAsnLysMetLysValAspSerSerPheLysMetAsn--- 276
778 GACATATACAATTAACCTTTTCATGATATATGAAATCTTACTGGAATATGAATG 837
277 SerTyrGlnLysValIlePhe-----AsnAsn----- 285
838 ACACGTTTTCAGTACTACCAACCAATATCAAGATATTAATATCTTATACACATTATCAT 897
286 -----AspTyrSerIleSerCysVal----- 292
898 TTCATGATGATTTTATGACATATTAATCATCTTCGCGGTTTAATATG 957
293 -----MetAsnSerLysThrValIleProSerLeuTyrAsnLysSerIleAsnPhe 309
958 -----TATTAACACCAATATC-----ATAACCAAACTAATGATGAGCCT 996
310 GlnLeuLysAlaAsnPheAsnValAspHisGlnLeuIlePheLysLeuIleSerLysAsp 329
997 TGTTTTTCATTTGACATTAAT-----TCGAGTATTCCTTATGT- 1034
330 LeuTyrAsnMetLysIleAsnGlnLysLeuValPheLeuAsnPheSerAspTyrProPhePhe 349
1035 ---GATGTATCATGAAAAATTCCAACATGGTTTATCTTTACGAACTCATTCAGAAC 1091
350 IleLysLeuGlnSerArgAsnLeuSerCysValIleLysLysAsnTyrIlePheLysLeu 369
1092 AAC-----GTAAATCCCACTCTTTTATAGTATGATGACAA 1124
370 LysSerPheAspGlnValLeuLysGlnTyrIleAsnAsnTyrPhePhe-SerLeuLysAs 389
1125 TATTTTTCATTTATTAAGAT-----GATTAAGATGATTT 1160
388 nIlePheThrLeuGlnAspLeuProIlePheIleAspIleGlnGlnArgGlnAspLe 409
1161 TAACAGATTTTAAATTAATTAATTAATCAACGTGTATTAACGTAATGTAATAA 1220
409 uAsnAsnIlePheLeuLysLysIleAsnPhePheProIleLysGlnLys-----LysPhe 427
1221 CTAATAT-----AATGATATGATTAAGTATATATCAATTAACAATTAATTAAGAT 1271

427 eTyrLysLysValIleHisProGlnAspTyrIleLysTyrAsnGlnTyrIleLeuLysLe 447
1272 GATTCA-----GACATTACGGGT----- 1290
447 uIleGlnGlnIleAsnIleThrGlnLysSerAspArgHisThrHisTyrValHisIlePr 467
1291 -----ATTGATTGCAATGCA 1304
467 oLysIleAspLeuTyrAlaAsnIleMetLysLysLysLeuSerIleLeuGlnValLeuTyr 487
1305 TATACGGTGAATTCGTTTCTT----- 1326
487 rTyrLysAsnPheAsnPheIleGlnThrProGlnIleAsnLeuLeuLeuGlnLysAsnLys 507
1327 -----AATATGAATGTGAATACCTTTCA 1349
507 sLeuTyrLeuArgGlnSerLeuGlnLysLysTyrAsnIleTyrSerSerIleTyrAlaAs 527
1350 TGCACGTGATATTATTTTCAAACTATTTTATTAACACAAAGTAACTTAAACAA 1409
527 nAsnLeuAspTyr-----PhePheProLysLeuGlnGlnArgMetGlnAlaLeu 543
1410 AATCAATATGACATTCACCTTACGACTATCATCTAGTAT-----GA 1451
543 -----SValAsnPheLeuGlnLysAsnLysPheProIleIleSerSerLysIleLeuAlaArgAs 563
1452 TATCAAC--GAACACCCATCTCAATGAAGAGGTATGTATCTAAAGTCGTTTAA 1508
563 pLeuAsnThrPAsnAsnIleTyrPheLysAsnIleLysValLeuThrGlnLysAsnIleAs 583
1509 TGCATTTATAGGC-----ATACCTGCATTAACCTTCATTTAATCTTATCCG 1556
583 nAsnThrPheSerGlnLysKetLeuIleTyrAlaAsnLysIleHisPheTyrLysPheTyr 603
1557 TTTAGAT-----GATTAACAATGAATATACAT----- 1584
603 rIleAsnThrLeuHisIleGlnThrTyrSerAsnAsnHisLysGlnAsnPheSerPheLe 623
1585 -----ATCATTAACGGTTTACAAACACTGAA-- 1611
623 uLeuLysSerAsnArgLeuHisIleAsnLeuIleIleAsnGlnAlaPheAsnAsnLysThr 643
1612 -----GCTAATATATATCTCTACATTT-----GTCAAC 1640
643 rGlnHisThrHisGlnPhePheLysLysIleAsnIleArgThrPheThrGlnValThr 663
1641 ATCAGCTTCATGTATTAACCTTATTTGTTCCCTTCCAACTACTTAACGGAAGTGAATTGA 1700
663 rAlaLysLys----- 666
1701 CGACATTTTATTT---TATTCGCACT-----GATAGTTTGTATATGAATC 1745
667 -AsnAsnPheIleHisTyrTyrAspSerHisAsnSerIleThrAsnPheTyrGlnLysSe 686
1746 CGTTGTAAACCTTATTAAGAACCCAGTTTATTCGACCCAGTACCTTACGTAAATGGCA 1805
686 rIleLysLysArgAsnLysPheSerSerPheLeuTyrAsnValLysMetSerPhePheAs 706
1806 TATTAAGAACCAACAGATAGATAGATGTTGTCTGAACATCAAGAAATATGATATGA 1865
706 nLeuPheAsn-----ArgSerPheIleSerPheGlnSerLysLeuSer----- 720
1866 AGTGAATGGAAGATTAATTT-----GCTTCGCTGCT 1899
721 -IleAsnAlaLysLeuLysLeuIleLeuGlnLysMetIleSerAspGlnAlaIlePheLe 740
1900 -----ATACCGAAAAACCCCTTGATTAACAGCT 1928
740 uLysGlnLysAsnAsnThrLysLeuGlnLysLysIleAsnLysLysIlePheIleGlnAsnIle 760
1929 CGATTTT-----GAACCTTTGTACGTGAACAAATCTTTGACGGTGCATTAAT 1976
760 eAspPhePheLysIleSerMetAsnLeuIleLysAsnAspPheLysSerLysThrIleIle 780

```

QY 1977 TGAAACAAATTAAGTATCTATATGACGAAGTACATATATGATATACCTCTAAAC 2036
    ||||| ||| |||
Db 780 elyslyasnllylsleuserasnasnlysasnllephcglytyleu----- 796
QY 2037 TGAATTTGTAATGTCGATATATATGATATATTTTACTGATGAACCTTAATGAAACG 2096
    ||||| ||| |||
Db 797 -----Asnilleaspletyr-----Asnlylsyasnllelysgl 809
QY 2097 TGAATTTATA 2106
    ||||| |||
Db 809 yglupheile 812

RESULT 12
EXEN_CLOBO STANDARD; PRT: 1162 AA.
AC P46082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Botulinum neurotoxin type E, nontoxic component.
GN ENT-120.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
CX Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maslike;
RX MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
  component of Clostridium botulinum type E progenitor toxin."
  J. Gen. Microbiol. 139:79-86(1993).
  1- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
  TOXICITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC FMBL: D12697; BAA02194.1;
DR InterPro: IPR000385; Bontoxilysin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDDAE0F39D CRC64;

Alignment Scores:
Pred. No.: 6,48e-05 Length: 1162
Score: 198.00 Matches: 185
Percent Similarity: 32.40% Conservatve: 140
Best Local Similarity: 18.44% Mismatches: 276
Query Match: 4.86% Indels: 402
Db: 1 Gaps: 51

US-09-727-892a-2 (1-2286) x EXEN_CLOBO (1-1162)
QY 103 AAACCAACCAATTAATAAAGCTTACTTATCTGATGCAATGTTGGTTAATGCTTAT 162
    ||||| ||||| ||||| |||||
Db 203 LysPrcAsnAspAsnleuSnlleProtyrAlaGluAaG---AsnGluPhcAsnSerleu 221
QY 163 GAAATTTGATGTTGAAGTATTTCCGAGTTTGCATCTTTTATGACGCAATTTATACGAT 222
    ||| ||| ||| |||
Db 222 Glu-----TyrSerGlu 225
QY 223 GTGAAGAAAGCGTATCATCATCAAAATCAAAAGACATATTATCATGATGTCACATAC 282
    ||| ||| ||| |||

```

```

Db 226 LeuAsnMetIleAspPheLeuIleSerGlyIleAspTyrIleLeu-----Asn 243
QY 283 TGTATTAATAC-----GATATCATTTTACTTAAGACACATGCGTATTTT 333
    ||| ||| ||| |||
Db 244 ThrAsnProtyrTrpPheIleAspLysTyrPheIle-----AspThrSerIleAsnPh 261
QY 334 GATATATTATACACGGAAATATATATTAAATCTGCAGAAAGAAATGACACACATTA 393
    ||| ||| ||| |||
Db 262 GluLysTyr-----LysAsnAspTyrGluIle 270
QY 394 AAATGAAAGAGGCTACTATTTTTACCCAAATCAAAATGTAATTTTGAAGAAACGCTT 453
    ||||| ||||| ||||| |||||
Db 271 LysIleLysAsnAsnAsnTyrIleAlaAsnSerIleLysLeuTyrLeuGluGlnLysPhe 290
QY 454 AATCTTCATC-----AATTAAGCTTTAAACAATGTTTAAATGCTTT--- 498
    ||| ||||| ||||| |||||
Db 291 LysIleAsnValLysAspIleTrpGluLeuAsnLeuSerTyrPheSerLysGluPheGln 310
QY 499 -----AAATTTAAT 507
    ||| |||
Db 311 IleMetMetProGluArgTyrAsnAsnAlaLeuAsnHisTyrTyrArgLysGluPheTyr 330
QY 508 ATTATTGATTAATCTT-----ATGAAAC 531
    ||||| ||| ||| |||
Db 331 ValIleAspTyrPheLysAsnTyrAsnIleAsnGlyPheLysAsnGlyGlnIleLysThr 350
QY 532 AATACATCAATTCACACATTAAGTACAAATTAATCTT----- 567
    ||||| ||||| ||||| |||||
Db 351 LysLeuProLeuSerLysTyrAsnLysGluIleIleAsnLysProGluLeuIleValAsn 370
QY 568 -----GATGT- 573
    ||| |||
Db 371 LeuIleAsnGlnAsnAsnThrValLeuMetLysSerAsnIleTyrGlyAspGlyLeuLys 390
QY 574 -----GGTATTTAACAGATCAACATCACTTAACAGATTT 609
    ||||| ||| |||
Db 391 GlyThrValAspAsnPhcTyrSerAsnTyrIleIleProtyrAsnLeuAsnTyrGluHis 410
QY 610 AATTATACGATTTTGTATTAAGATAT-----GAT 639
    ||| ||||| ||||| |||||
Db 411 SerIleAsnTyrPheTyrLeuAspAsnValAsnIleGluGlnIleGluLysIleProPro 430
QY 640 ATGATGATGATGGAAGCCTATGACTAT----- 666
    ||||| ||||| ||||| |||||
Db 431 IleAsnAspGluAspIleTyrProtyrArgLysAsnAlaAspThrPheIleProValTyr 450
QY 667 -----GCTGTGAATGTGTTGCAAAACCTCACACCTGCAACACTTACATTCAT 717
    ||||| ||||| ||||| |||||
Db 451 AsnIleThrLysAlaLysGluIleAsnThrThrProLeuProValAsnTyrLeuGln 470
QY 718 AATGACGTGATTAATATGATGTGTCATATTCATTAAGTATATATTTCCAAATTT 777
    ||| ||| ||| |||
Db 471 AlaGlnMet-----IleAspSerAsnAspIleAsnLeuSerSer 483
    ||| ||| ||| |||
QY 778 GACTATACAAATTAATTAATTTTCATTTG---AATATTAAGCATCTTACTTAATGAAG 834
    ||||| ||||| ||||| |||||
Db 484 AspPheLeuLysValIleSerSerLysGlySerLeuValTyrSerPheLeuAsnAsnThr 503
QY 835 ATGACACGTTTCAGTACTACCAACCAATATCAACATATTAATATATCTTATPACACATTA 894
    ||| ||| ||| |||
Db 504 MetAspTyrLeuGluPheIleLysTyrAspLysProIleAspThr----- 518
    ||| ||| ||| |||
QY 895 CATTTCCATGATTAATTAATTTTATGACTATTAATTAATCAATCTTCGCTGCT 945
    ||| ||| ||| |||
Db 519 -----AspLysLysTyrTyrLysTyrPheLysAlaIlePheAlaGsnTyrSerLeu 535
    ||| ||||| ||||| |||||
QY 945 ----- 945
    ||| ||| ||| |||
Db 536 AspIleThrGluThrGlnGluIleSerAsnGlnPheGlyAspThrLysIleIleProTyr 555
    ||||| ||||| ||||| |||||
QY 946 -----GGTTAAATATGATATTAACACCAATATCAACAAATTAATGATGAG--- 993
    ||||| ||||| ||||| |||||
Db 556 IleGlyThrGlnAlaLeuAsnIleLeuAsnThrAsn-----AsnSerPheValGluGluPhe 573
    ||||| ||||| ||||| |||||

```

```

QY 994 -----CCTGTTT 1002
Db 574 LysAsnLeuGlyProIleSerLeuIleAsnLysGluAsnIleThrIleProLysIle 593
QY 1003 TCATATGAC----- 1011
Db 594 LysIleAspGluIleProSerSerMetLeuAsnPheserPheLysAspLeuSerGluAsn 613
QY 1012 -----ATCATTCGAGTTATCCCTTATGATGATCATCTGAAAAATT 1053
Db 614 LeuPheAsnIleTyrcysLysAsnAsnPhetyrLeuLysLysIleTyrrAsnPhelu 633
QY 1054 CCACATGCTTATACCTTTACGACACTATTCAGAACCAAGCTTAATCCCTACTTTTAA 1113
Db 634 AspGlnTrp-----TyrPheGlnTyrrTyrrSerGln----- 643
QY 1114 GATGATGACATTAATTTTTCATTATAT--AAGATTGATTAAGATGATTTAAGATGAT 1170
Db 644 -----TyrPheAspLeuIleCysMetAlaSerLysSerValIleuAlaGlnGlu 659
QY 1171 TTATTAAT--AAATTAATACACGCTGATTAAGCAATGATGTAATAATCTATAT 1227
Db 660 LysLeuIleLysLysLeuIleGlnLysGlnLeuArgTyrrLeuMetGluAsnSerAsnIle 679
QY 1228 AATGATATGATTACGTTAATATCAAT-----ACAAATACATTAGAATGATTCAA 1278
Db 680 SerSerThrAsnLeuIleLeuIleAsnLeuThrThrAsnThrLeuArgAspIleSer 699
QY 1279 GAC-----ATTACGGGTATGAT-----TGC 1299
Db 700 AsnGlnSerGlnIleAlaIleAsnAsnIleAspLysPhePheAsnAsnAlaAlaMetCys 719
QY 1300 -----ATGCATATACGTTAATTCGTTGTT----- 1326
Db 720 ValPheGluAsnAsnIleTyrrProLysPheThrSerPheMetGluGlnCysIleLysAsn 739
QY 1327 -----ATATATGAAATGGAATAC 1344
Db 740 IleAsnLysSerThrLysGluPheIleLeuLysCysThrAsnIleAsnGlnThrGluLys 759
QY 1345 TTTCATGACAGTGTATATTTTCAAACTATTATTATTAACACAAA----- 1392
Db 760 SerHis-----LeuIleMetGlnAsnSerPheSerAsnLeuAspPheAspPheLeu 776
QY 1393 -----GGTAAGTTAAAAACAAATCAATG----- 1419
Db 777 AspIleGlnAsnMetLysAsnLeuPheAsnLeuTyrrThrGluLeuLeuIleLysGluGln 796
QY 1420 ACATCACCTTAC-----GACTATCACATTAAGTAT 1449
Db 797 ThrSerProTyrrGluLeuSerLeuTyrrAlaPheGlnGlnGlnAspAsnValIleGly 816
QY 1450 GATATCAC-----GAACACCCA----- 1467
Db 817 AspThrSerGlyLysAsnThrLeuValGluTyrrProLysAspIleGlyLeuValTyrrGly 836
QY 1468 TATCAATAGAGAGCTTATCTATCT-----AAAGTCGTTTAAATGATTA 1515
Db 837 IleAsnAsnAsnAlaIleHisLeuThrGlyAlaAsnGlnAsnIleLysPheThrAsnAsp 836
QY 1516 TATGGCATACCTGCATTCATTCATCAATTTAATCTATTC----- 1554
Db 857 TyrPheGluAsnGlyLeuThrAsnAsnPheserIleTyrrPheTrpLeuArgAsnLeuLys 876
QY 1554 ----- 1554
Db 877 GlnAsnThrIleLysSerLysLeuIleGlySerLysGluAspAsnCysGlyTrpGluIle 896
QY 1555 CGTTAGATGATACATGAATGATACATATCATACGTTCAAAACACAGTGAAGT 1614
Db 897 TyrPheGluAsnAspGlyLeuValPheAsnIleIleAspSerAsnGlyAsn---GluLys 915

```

```

QY 1615 AATATATATTCCTACATTTGTCACATCAGCTTCATGTTATACCTATGTCCTTTC 1674
Db 916 AsnIleTyrrLeuSerAsn---IleSerAsnLysSerTrpHisTyrrIleValIleSerIle 934
QY 1675 CAATACCTAACGGAAGTGAAT-----GACGACATTTATTTATTCGATACGTAT 1728
Db 935 AsnArgLeuLysAspGlnLeuLeuIlePheIleAspAsnIleLeuValAlaAsnGluAsp 954
QY 1729 AGTTTATATAGAAATCCCTGTTAAACCTTATGAACCCAGCTTATTCAGCCGATA 1788
Db 955 -----IleLysGluIleLeuAsnIleTyrrSerSerAspIleIle 967
QY 1789 GCCTTAGGTAAATGGGATATGTAAGAAACGACAGATAGATAGATGTTGTACTGATCAT 1848
Db 968 SerLeuLeuSer---AspAsnAsnAsnValTyrrIleGluGlyLeuSerValLeuAsnLys 986
QY 1849 AAGAAATATGCAATGACATG-----AATGGAAGAT 1881
Db 987 ThrIleAsnSerAsnGlnIleLeuThrAspTyrrPheSerAspLeuAsnAsnSerTyrrIle 1006
QY 1882 AAATTCCTTCCTGCTGCTATACCGAAGAAACGCCCTTGATACAGGCTC----- 1929
Db 1007 Arg-----AsnPheAspGluGlnIleLeuGlnTyrrAsn 1017
QY 1930 ---GATTGGAACCTTTGACGTGACAAATTCCTTGACGGTCCATT-----ATTGA 1980
Db 1018 ArgThrTyrrGluLeuPhe---AsnTyrrValPheProGlnIleAlaIleAsnLysIleGlu 1036
QY 1981 AACAATTAAGATCTAT-----AATGACAGAGTACATATGATATATCCGCTCT 2031
Db 1037 GlnAsnAsnAsnIleTyrrLeuSerIleAsnAsnGluAsnAsnLeuAsnPhelysProLeu 1056
QY 2032 AAAATGAAATGTATGATGCTATGATATGATATGATATTTACT-----GATGAAT 2085
Db 1057 LysPheLysLeuLeuAsnThrAsnProAsnLysGlnTyrrValGlnLysTyrrAspGluVal 1076
QY 2086 AATATGAAACGTATTTATTTATTAAGACGCTAGAGAAATTC----- 2130
Db 1077 -----IlePheSerValLeuAspGlyThrGlnLysTyrrLeuAspIleSerThr 1092
QY 2131 -----GACCATGATCAATTTGATGATATCTTATATTTGAA 2166
Db 1093 ThrAsnAsnArgIleGlnLeuValAlaAspAsnLysAsnAsnAlaGlnIlePheIleIleAsn 1112
QY 2167 AGTGACATC 2175
Db 1113 AsnAspIle 1115

RESULT 13
DPOM.PODAN STANDARD: PRT: 1197 AA.
AC 001529;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Podospora anserina.
OG Mitochondrion.
OC Plasmid pAL2-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriales; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL2;
RX MEDLINE=93113721; PubMed=1473181;
RA Hermans J., Osiewacz H.D.;
RT "The linear mitochondrial plasmid pAL2-1 of a long-lived Podospora
RT anserina mutant is an invertion encoding a DNA and RNA polymerase.";
RL Curr. Genet. 22:491-500(1992).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER

```


Db 1260 nlysaasnval-----SerTyrIleHisasnaspGlnGlyLysHI 1273
 QY 475 -TTACATGTTTAAATGGTTTAAATTAATTAATGATGATGATTAATGACAAACCA 533
 Db 1273 sleuSerleuGlnMetCyslaGlnGlnHisIleThrIleAsnThrAsnaspAs 1293
 QY 534 TACATCATTCGACACATAGGTAGAGAAATTAAGTGGTGGTATTAAACAGATCA 593
 Db 1293 nThrTyrValGlnGlnIleGlnMetLysGlnLeuSerLysasnLysThrLysGlnLysGln 1313
 QY 594 ACTTAACAGATTTAATATATGATTTTGTAT-----AAAGATATGATATGATGA 647
 Db 1313 n-----SerPheLysGlnThrIleArgaspMetHisGlnaspSerGlnGlnLys 1330
 QY 648 TAGTGAACCTATGATAGCTAGCTGAAATGTTTGCAGAACTGCACCTGCAGACATTA 707
 Db 1330 taanLysPheIleThrLysLysAlaLysPhePheIleGlnLysLysGlnLysMetHis 1350
 QY 708 ATACATTCATATGACGTGATTAATAGTAGTGTCCATATTCATTAATAGCATATA-- 765
 Db 1350 sGlnCysAsnaspAsp-----IleGlnTyrAsnThrGln 1362
 QY 766 -TTTCCAAATTTTGCATATACAAATTAACATTTTCATGATATATGATATGATCTTACT 824
 Db 1362 nTyrAspAsnIleGlnTyrAsnaspIle-----SerCysAsnTyrIleLysSerGln-- 1379
 QY 825 GAATTAATGAATGACACGTTTTCAGTTTACACCAATTCAGATATTA-- 876
 Db 1380 -AsnLeuGlnAsnThrHisIleGlnValAsnaspLeuSerPheIleLysasnAsnVal 1399
 QY 877 ----ANATCTTATACATATCAT-----TTCATGATATGATATTTAT----- 918
 Db 1399 IileLeuProLysGlnLysThrHisSerIlePheHisPheValAsnAspTyrArgasnVal 1419
 QY 919 -----GACTATATTAATCATTTCTATCGTGGTGT---TTAA 953
 Db 1419 lvalGlnLysasnLeuMetaspLysLysIlePheLeuAsnaspSerGlnLysAs 1439
 QY 954 TATGTAATACCAATTAACATTAACAACTAATGATGACCTGTTGTTTCTATTGACAT 1013
 Db 1439 nvalValGlnSerLysTyr---AsnArgMetSerLysasnLeuLysLysIleGlnL 1458
 QY 1014 CAATTCGAGTTATCTTATGTGATGTATCATGAAAAAATTCACACGTGTTTACTTTTA 1073
 Db 1458 eileasnaspIle---TyrArgasnGlnLysLysLysLeuAsnArgTyr----- 1473
 QY 1074 CGAACACTATTCAGAACCAAGCTTAATCCCTACTTTTATAGAT---GATGACATATTT 1130
 Db 1474 -----LysThrLysMetaspAspAspAsnTyrTr 1484
 QY 1131 TTCATTAATATAGATGATGAAGATGTATTTAAGATGATTTAATTAATTAATTC 1190
 Db 1484 pSer-----SerAspAspSerIleIleAlaLysLysIleLysLysAs 1501
 QY 1191 ACCTGATTAACGCAATGATGTTGTAATATCATATATATGATTAATGATTAGCTTAAT 1250
 Db 1501 nlysaGlnLysArgLys-----TyrHisProLysGlnGlnGlnAsnHis 1516
 QY 1251 CAATTAACATTAACATTAACATGATGATGACAGATTAACGCTGATGATGATGATGAT 1310
 Db 1516 eaSprArgasnAsnTyrLysMetIleThrArgasnAsnasp-----As 1532
 QY 1311 TGTTAATTCGTTTGTATATATGAAATGGAATACCTTTCATGACGCGATATATTTC 1370
 Db 1532 nAsnAsnAsn-----As 1536
 QY 1371 AATCTATTTTAAATACAAAGGTAAAGTTAAACAAATCAATATGATGATCACTCACT 1430
 Db 1536 nAsnAspAsnAsnAsnAspAsnAsnAsnAsnAsnAsnSerAsnAsnAsnTyr 1556
 QY 1431 CGACATATCACTATGATGATATCAACGACACCACTCACTCAATGACGAGTTATGTT 1490
 Db 1556 tGlyTyrAsnLeuHisIleAspAspValAsnaspLeuGlnValThrAsnTyrAsnThrAsn 1576

QY 1491 ATCTAAGTCGTTTAAATGGATTAAT-----GGCATACCTGCATTAACGTTTC 1538
 Db 1576 eTyrProAsnaspCysasnGlnLysIleTyrGlnLysGlnLysThrAsnaspAsnGlnLeuThrTh 1596
 QY 1539 ACATTTTAATTAATTCGTTTATGATGATTAACAT-----GACTATATACATAT 1586
 Db 1596 rAsnSerAsnMet-----CysAspLysAsnaspAspSerAspGlnPheAsnAs 1614
 QY 1587 CATTAACGCT-----TACAAAAACACTGAACGTAATAT 1619
 Db 1614 nIleasnGlnAsnaspLeuLeuTyrAspAsnLysTyrTyrArgGlnIlePheLysasnVal 1634
 QY 1620 ATTA-----TTCTACATTTGTCACATCAGCTTCATGTAATTAATTTGCT 1667
 Db 1634 IileGlnPheValSerValPheGlnTyrValGlnSerTyrLysGlnHisTyrIleLeuPhe 1654
 QY 1668 TCCTTTTC-----CAATACGTAACGGAAGTAAT 1697
 Db 1654 eProTyrGlnIleIleLysTyrPheThrSerPheLeuLeuGlnLysThrGlnIleLeuPhe 1674
 QY 1698 TGACGACAAATTTATTTAATGCGATACGAT----- 1728
 Db 1674 oThrAsnIlePheLeuHisThrLysLeuSerLysLysGlnLysProThrHisGlnLysAs 1694
 QY 1729 -----AGTTGTATTAAGAANCCTGTTAAACCTTATGAACCC----- 1770
 Db 1694 nThrGlnLysMetLysLysIleTyrIleGlnGlnLysLysThrPheLeuPheIleLysAlaI 1714
 QY 1771 -----AGTTTATTCGACCCGATACCTTAAGTA-- 1800
 Db 1774 easnIleTyrLysTyrPheSerPheLysSerIleLeuLeuLysLysLysAspTyr 1724
 QY 1801 -----TGCATATTAAGAAACGACATAGATTAACATGTTTACTGAAT----- 1845
 Db 1744 rPheAsnTyrIleIleLysasnTyrAspIleSerHisArgTyrIleIleHisAspTyrSer 1754
 QY 1846 -----CATAGAATATATGC 1859
 Db 1744 rPheIleAsnLeuLysGlnLeuTyrLeuPheIlePheHisAsnIleLysTyrPheLys 1774
 QY 1860 ATAT-----GAAGTGAATGAAAGATTAAATTCCTTCTGCTGTAATAC 1904
 Db 1774 sTyrIleSerThrProGlnAspAlaValAlaGlySerIleSerAlaGlnSerIleGlyGlnP 1794
 QY 1905 GAAACACCTTTGATACACGCGGATTTGAACCTTTAGCTGAACAATTCCTTGA 1964
 Db 1794 o-----GlyThrGlnMetThrLeuLysThrPhe-----HisPheAl 1806
 QY 1965 CGGTGCC-----ATTATGAAA 1982
 Db 1806 agLysAlaSerMetAsnValThrLeuGlnValProArgIleLysGlnIleIleAsnAl 1826
 QY 1983 CAATTAAGTATAC-----TATTA 2000
 Db 1826 aserAsnSerIleGlnThrProIleLeuAsnIleProLeuGlnValAsnaspAsnTyrAs 1846
 QY 2001 TGACGACGTCATATATGATATATCCGCTTAACATGAAAT-----GTATGCTGTA 2054
 Db 1846 nPheAlaLeuMetIleLysSerLysLeuGlnLysThrThrIleArgaspIleCysMetLys 1866
 QY 2055 TGATATATGATTAAT-----TTTACTGATGACT 2084
 Db 1866 rIleLysGlnLysPyrThrSerArgGlnValPheLeuSerValLysPheAsnGlnGlnLys 1886
 QY 2085 TAATATGAACGTAATTTATATTAATAAGACGTAGAAATTTTCACCATATGATCAT 2144
 Db 1886 u-----IleGlnLysLeuPheLeu-----AsnIleAsnAlaTyrAsnI 1899
 QY 2145 TGATGATTTCTTATATATGAAGTGAACATCGGTTCATTTTCACTTAACGATTAATTC 2204
 Db 1899 eLysAspIleIleLeuLysGlnSerHisIleAsnLysIleLysIleAsn-----Ly 1916

QY 2205 AGTTGAAGCTTCAGTACATACAAATCTGATTGCATATA---TTAAACGTGACAT-- 2259
 Db 1916 sileHsILsAsnValILeNsILsYrILsLeuHsILsSerLeuILsAsnSpGLuPh 1936
 QY 2260 -----GATGAATTAATAAAAGCG 2277
 Db 1936 eilePhegHleMleGluSerLeuILsYsGLy 1947
 RESULT 15
 TOXA_CLODI
 ID TOXA_CLODI STANDARD: PRT; 2710 AA.
 AC P16154;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Toxin A.
 GN TOXA OR TCDA.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1496;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RX MEDLINE=90221894; PubMed=2109310;
 RA Sauerborn M., von Eichel-Streiber C.;
 RT "Nucleotide sequence of Clostridium difficile toxin A.";
 RL Nucleic Acids Res. 18:1629-1630(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RX MEDLINE=90129305; PubMed=2105276;
 RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyerly D.M.,
 RA Wilkins T.W., Johnson J.L.;
 RT "Molecular characterization of the Clostridium difficile toxin A
 RT gene.";
 RL Infect. Immun. 58:480-488(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VPI 10463;
 RA von Eichel-Streiber C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
 CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
 CC DIFFERENT OLIGOPEPTIDES.
 CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
 CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
 CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
 CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X51797; CAA36094.1;
 CC DR EMBL: M30307; AAA23283.1;
 CC DR EMBL: X92982; CAA63564.1;
 CC DR PIR: S08638; S08638.
 CC DR InterPro: IPR002479; CM_binding.
 CC DR Pfam: PF01473; CM_binding_1; 31.
 CC KW Toxin; Enterotoxin.
 CC SQ SEQUENCE 2710 AA; 308052 MW; 0A6E52CE84C14421 CRC64;
 Alignment Scores:
 Pred. No.: 0.000445 Length: 2710
 Score: 183.50 Matches: 172
 Percent Similarity: 34.66% Conservative: 133
 Best Local Similarity: 19.55% Mismatches: 270
 Query Match: 4.51% Indels: 305

DB: 1 Gaps: 46
 US-09-727-892a-2 (1-2286) x TOXA_CLODI (1-2710)
 QY 175 GAAGTATTTCCGAGT-----TTGCAATCTTTTATGACCGCATTTATAGC 219
 Db 1258 AspleuTyProclYsPheTyTrPaGpHeTyAlaPhePheSpTyAlaIleHr 1277
 QY 220 TATGTGAAA-----AGAGCTGATACATACCAAAAATCAAAACAGAT----- 261
 Db 1278 ThrLeuSproValIryGluSpThrAsnILeYsILeYsLeuSpsThraG 1297
 QY 262 ---ATATCATGATTCACATTAACGTATTAATAGATATCATCTTTTACTTAAAGC 318
 Db 1298 AsnPhelIleKerProHrIleHrThrAsnGLU-----IleArgAsn 1311
 QY 319 ACCATGCGTTAT---TTGAT----- 336
 Db 1312 LysLeuSerTySerPheAspGlyAlaGlyThyTySerLeuDeuLeuSerTy 1331
 QY 337 -----ATATTTACGCGCAAAATATATTTAAATCTGCAGAGAA 378
 Db 1332 ProIleSerThrAsnLeuSerLysAspSpleuTrpIleAsnIleAspAsn 1351
 QY 379 AATGAACACACATTAATAATGAAGAAGGCTACTATT----- 414
 Db 1352 GluValArgGluIleSerIleGluAsnGlyThrIleYsLeuGlyLysLeuIleYsAsp 1371
 QY 415 -----TTAGCAAAATCAAAAGTAAAT----- 438
 Db 1372 ValLeuSerLysIleAspIleAsnLysAsnLysLeuIleGlyAsnGlnIleAsp 1391
 QY 439 -----TTAGAAAA 447
 Db 1392 PheSerGlyAspIleAspAsnLysAspArgTyIlePheLeuThcCysGluLeuSpsAsp 1411
 QY 448 CGGTGTAATCTTCATCATTTAGATT-----ACAATGTTTTAAT 492
 Db 1412 LysIleSerLeuIleIleGluIleAsnValAlaIleYsSerTySerLeuLeuSer 1431
 QY 493 GGTTTAAATTTAAATTTATTTGATCTTATGAAGAACCAATATCATCTTCACATTA 552
 Db 1432 GlyAspLysAsnTyIleuIleSerAsnLeuSerAsnThIleGluLysIleAsnThrLeu 1451
 QY 553 GGT-----AAGAATTAAGTGTGTTATTAAACAAATCAAA----- 594
 Db 1452 GlyLeuAspSerLysAsnIleAlaTyAsnTyThrAspGluSerAsnLysTyPhe 1471
 QY 595 -----CTTAACACATTTTATTTATAGATTTTGTATGAATGAATGATATGAT 645
 Db 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHsILsYsLysAspSerLysAsn 1491
 QY 646 GATAGTGAAGCCATGACTATGCTGTAATGTTTGGCAAACTCACACCTGACAACTT 705
 Db 1492 IleLeuGluPheTyAsnSpsThrLeuGluPhe-----AsnSerLys 1506
 QY 706 ACATATATTCATTAATACCGTATATATATAGGATGTGCAATTCATATAGTATATA 765
 Db 1507 AspPheIleAlaGluSpsIle-----TTGAAVal 1515
 QY 766 TTTCCAATTTTGCATTAACAATTAACA----- 795
 Db 1516 PheMetLysAspSpsIleAsnThrIleThrGlyLysTyTyValAspAsnSnrThraSp 1535
 QY 796 -----TTTCATGGAATATATAGAAATC----- 819
 Db 1536 LysSerLysAspPheSerLysLeuValSerLysAsnGlnValLysValAsnGlyLeu 1555
 QY 820 TACTGTAATTAAGAATGACAGCTTTTCAGTTACTACCAATATCAAGAT---ATTAA 876
 Db 1556 TyIleAsnGlu-----SerValTySerTyLeuAspPheValLys 1570
 QY 877 ATATCTTATACACATATATCATTTCCATGAT---ATCAAT---TTTATGACATATATTA 930

Db 1571 AsnSerAspGlyHisHisAsnThrSerAsnPhenMetAsnLeuPheLeuAspAsnLeu---- 1589
 QY 931 TCATTTATCGTGGTGGTAAATATGATTAACACCAATACATACAACTAATGAT 990
 Db 1590 SerPheTrpLys-----LeupheGlyPheGluAsnIleAsnPhenValIleAsp 1605
 QY 991 GACCCGTTGTTTCTATGACATCAATTCGAGTTATCCCTATGATGATCATGAAAA 1050
 Db 1606 Lys---TyrPheThrLeuValGlyLysThrAsnLeuGlyTyrVal----- 1619
 QY 1051 ATTCACAATGGTTATCTTTTACGACACTATCAGAACCAACGTTAATCCCTCTTT 1110
 Db 1620 -----GluPhe 1621
 QY 1111 TTGATGATGACAAAT-----TATTTTCTATADATAGATGATGAAGAT 1155
 Db 1622 IleCysAspAsnAsnLysAsnIleAspIleTyrPheGlyGluTrpLysThrSerSer 1641
 QY 1156 GATTTTAACGATTTATTAATTAATAATTAATCAACGCTGATTA-----CGT 1203
 Db 1642 -----LysSerThrIlePheSerGlyAsnGlyArg 1651
 QY 1204 CAATATGTTGAATATCTATATATATGATGATGATGATGATGATGATGATGATGAT 1263
 Db 1652 AsnValValValGluProIleTyrAsnProAspTrpGlyGluAspIleSerThrSerLeu 1671
 QY 1264 TTAGATGATATTCAGACATTCAGCGGTATGAT----- 1296
 Db 1672 AspPheSerTyrGluProLeuTyrGlyIleAspArgTyrIleAsnLysValLeuIleAla 1691
 QY 1297 -----TGCAATGCATATACGTTGTTAATGCTGTTGTTATATATGAAGT 1338
 Db 1692 ProAspLeuTyrThrSerLeuIleAsnIleAsnThrAsn-----TyrTyrSerAsn 1708
 QY 1339 GAATACTTTCATGACGATGATATTTTTCACAAATATTTATTAATAACACAAAGTAG 1398
 Db 1709 GluTyrTyrProGluIleIleValIleAsnProAsnThrPheHisLys----- 1724
 QY 1399 TTAATAAACAAATCAATATG---ACATCACTTACGATACGATACATCACTGATGAT 1452
 Db 1725 ---LysValAsnIleAsnLeuAspSerSerPheGluTyrLysTrpSerThrGluGly 1743
 QY 1453 -----ATCAGCAGACCCCATCTCAATGAGAGGATGATGATCTAA 1497
 Db 1744 SerAspPheIleLeuValArgTyrLeuGluSerAsnLysLysIle---LeuGluLys 1762
 QY 1498 GTGCTTTTAAAGATATATGACATACCTGCATACGTTACGATTTTAACTTATTCGT 1557
 Db 1763 IleArgIleLys-----GlyIleLeuSerAsnThrGluSerPheAsnLysMetSer 1779
 QY 1558 TTGATGATACATCAATATACATATATCAATTAACGTTACAAAAACACCTGAACGTAAT 1617
 Db 1780 IleAsp-----PheLysAspIleLysLysLeuSerLeuGly 1791
 QY 1618 ATATATCTCTACATTTGTCACATACGTCATGCTGATATTAATTTGTTCCCTTCCA 1677
 Db 1792 TyrIleMetSerAsnPhenLysSerPheAsnSerGluAsnGluLeu-----AspArgAsp 1809
 QY 1678 TACTTAACGAAAGTGAATGACACAAATTTATTTATTCGATACGATGATGATGAT 1737
 Db 1810 HisLeuGlyPheLysIleIleAspAsnLysThrTyrTyrAspGluAspSerLysLeu 1829
 QY 1738 ATGAATCCGTTGTTAAACCCCTTATGACCCCACTTA-----TTGACCCCGATAGCC 1791
 Db 1830 ValLysGlyLeuIleAsn-----IleAsnAsnSerLeuPheTyrPheAspPro----- 1845
 QY 1792 TTAGTAAATGGATATTGAAGAAGACAGATAGATAGATGTTGTACTGATCATAG 1851
 Db 1846 -----IleGluPheAsnLeuValThrGlyTyrGluThrIleAsnGlyLys 1860
 QY 1852 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1902

Db 1861 LysTyrTyrPheAspIleAsnThrGlyAlaIleLeuThrSerTyrLysIleIleAsnGly 1880
 QY 1902 ----- 1902
 Db 1861 LysHisPheTyrPheAsnAsnAspGlyValMetGluLeuGlyValPheLysGlyProAsp 1900
 QY 1903 -----CCGAAAACGCCCTTGATACACACGCTGATTTTGAAACCTTT 1944
 Db 1901 GlyPheGluTyrPheAlaProAlaAsnThrGluAsnAsnAsnIleGluGlyAlaIle 1920
 QY 1945 GFA-----CGTGAACATTCCTTGACGGT----- 1968
 Db 1921 ValTyrGluSerLysPheLeuThrLeuAsnGlyLysLysTyrTyrPheAspAsnAsnSer 1940
 QY 1969 -----GCCATTTATGAACACATTAAGTATCTATTAATGACCAAGT 2010
 Db 1941 LysAlaValThrGlyTyrPheArgIleIleAsnAsnGluLysTyrTyrPheAsnProAsnAsn 1960
 QY 2011 ACAATATGATATATCCGCTTAACAACTGAATTTGTATGTGTAATGATATGATGATAT 2070
 Db 1961 AlaIleAlaIleVal-----GlyLeuGluValIleAspAsnAsnLysTyr-----Tyr 1976
 QY 2071 TTTACTGATGAA-----CTTATATGAACGTTGAA 2100
 Db 1977 PheAsnProAspThrAlaIleIleSerLysGlyTyrProIleThrValAsnGlySerArgTyr 1996
 QY 2101 TTTATATTAAGACGCTAGAGAAATTTTCACCATAGTCAATTTGATGAT-----ATT 2154
 Db 1997 TyrPheAspThrAspThrAlaIleAlaPheAsnGlyTyrLysThrIleAspGlyLysHis 2016
 QY 2155 CTTATATGAAAGTGAC-----ATCGGTTCAATTTTACTTAACGATTTATTT 2202
 Db 2017 PheTyrPheAspSerAspCysValValLysIleGlyValPheSerThrSerAsnGlyPhe 2036

Search completed: January 8, 2003, 17:07:09
 Job time : 66 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 65 Seconds
(without alignments)

14493.037 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 4070

Sequence: 1 atgggattactagaatgcatt.....taaaaaaggaacacgtttaa 2286

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ -n2p -model -DEV-xlh
-O=/cgn2_1/uspro.spool/US09727892/runat_06012003_151201_9300/app-query.fasta.1.2439
-DB=SPREMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blcsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-CURFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09727892.ecgn.1.1.83 &runat_06012003_151201_9300 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAT -LARGEOUTPUT -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_proteus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query
No. Score Match Length DB ID Description
1 257.5 6.3 960 5 Q25802 Q25802 plasmodium

2	252.5	6.2	925	8	Q9GBS2	Q9GBS2 flammulina
3	251	6.2	1321	8	Q95022	Q95022 tetrahymena
4	243.5	6.0	575	9	Q38545	Q38545 bacterioph
5	242	5.9	1121	5	Q96192	Q96192 plasmodium
6	237	5.8	826	12	Q9EMB8	Q9EMB8 amsecta moo
7	236.5	5.8	3973	5	Q96204	Q96204 plasmodium
8	230.5	5.7	1301	16	Q9XNG4	Q9XNG4 ciosstridium
9	230	5.7	1802	5	Q96170	Q96170 plasmodium
10	229.5	5.6	1182	5	Q96240	Q96240 plasmodium
11	225	5.5	1035	8	Q9XM75	Q9XM75 neurospora
12	220.5	5.4	571	8	Q78938	Q78938 agrocycbe ae
13	220.5	5.4	1822	5	Q9U0N4	Q9U0N4 plasmodium
14	219	5.4	578	9	Q96205	Q96205 bacterioph
15	218.5	5.4	872	12	Q9EMR4	Q9EMR4 amsecta moo
16	218	5.4	767	12	Q9EMU4	Q9EMU4 amsecta moo
17	218	5.4	979	5	Q9Y1K0	Q9Y1K0 plasmodium
18	218	5.4	1817	5	Q96253	Q96253 plasmodium
19	217.5	5.3	1410	5	Q97230	Q97230 plasmodium
20	215.5	5.3	900	8	Q94QR5	Q94QR5 pleurotus o
21	215.5	5.3	3844	5	Q94648	Q94648 plasmodium
22	215	5.3	1708	8	Q9GBS0	Q9GBS0 flammulina
23	214	5.3	1115	12	Q9QMH2	Q9QMH2 parvo-like
24	213.5	5.2	1019	10	Q98RM9	Q98RM9 guillardia
25	213	5.2	1306	12	Q9YVU0	Q9YVU0 melanoplus
26	212	5.2	807	12	Q9YVX4	Q9YVX4 melanoplus
27	212	5.2	1266	5	Q95258	Q95258 plasmodium
28	212	5.2	1928	5	Q9U0H2	Q9U0H2 plasmodium
29	211.5	5.2	1416	5	Q97309	Q97309 plasmodium
30	211	5.2	1267	2	Q9A1H1	Q9A1H1 carsonella
31	211	5.2	1417	5	Q97310	Q97310 plasmodium
32	210.5	5.2	2340	5	Q97298	Q97298 plasmodium
33	210	5.2	2567	5	Q9U0J6	Q9U0J6 plasmodium
34	209	5.1	1111	5	Q9U0K5	Q9U0K5 plasmodium
35	208.5	5.1	694	9	Q9FZR8	Q9FZR8 mycoplasma
36	208	5.1	1272	16	Q9BPR0	Q9BPR0 ureaplasma
37	207.5	5.1	1817	5	Q96206	Q96206 plasmodium
38	207.5	5.1	544	8	Q95529	Q95529 physarum po
39	207.5	5.1	547	8	Q35599	Q35599 physarum po
40	207.5	5.1	753	10	Q98S90	Q98S90 guillardia
41	206.5	5.1	1673	5	Q97332	Q97332 plasmodium
42	205.5	5.0	1304	5	Q9U0H4	Q9U0H4 plasmodium
43	205.5	5.0	2423	5	Q97393	Q97393 plasmodium
44	205.5	5.0	4961	5	Q97372	Q97372 plasmodium
45	205	5.0	1188	5	Q96143	Q96143 plasmodium

ALIGNMENTS

RESULT 1						
Q25802	PRELIMINARY;	PRT;	960 AA.			
ID Q25802;						
AC Q25802;						
DT 01-NOV-1996 (TEMBLrel. 01, Created)						
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)						
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)						
DE RPOD protein.						
GN RPOD.						
OS Plasmodium falciparum.						
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.						
OX NCBI_TaxID=5833;						
OX [1]						
RP SEQUENCE FROM N.A.						
RC STRAIN=C10;						
RX MEDLINE=96346169; Pubmed=8757284;						
RA Wilson R.J.M., Denny P.W., Preiser P.R., Rangachari K., Roberts K.,						
RA Roy A., Whyte A., Strath M., Moore D.J., Moore P.W., Williamson D.H.,						
RT "Complete gene map of the plastid-like DNA of the malaria parasite						
RT Plasmodium falciparum."						
RL J. Mol. Biol. 261:155-172(1996).						
DR EMBL; X95275; CA64574.1; -						
DR InterPro; IPR000722; RNA_pol_A.						
DR Pfam; PF00623; RNA_pol_A; 1.						
SEQUENCE 960 AA; 117987 MW; F41782D73607F35D CRC64;						

Alignment Scores:

Pred. No.:	7,68e-08	Length:	960
Score:	257.50	Matches:	189
Percent Similarity:	39.68%	Conservative:	136
Best Local Similarity:	23.08%	Mismatches:	298
Query Match:	6.33%	Indels:	197
DB:	5	Gaps:	46

US-09-727-892a-2 (1-2286) x Q25802 (1-960)

```

QY 126 TACTATTCGTGACATTCGTGGTTATGTTATGAAATTCATGATTCATTCCTC 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 TySerPheLeuTyRAsnTyRSeLeu-----AsnIleLys 52

QY 186 GAGTTTCGAAATCTTTTATGACCATTTATACGATGCAAAAGCGATACAAATC 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53 AspPheSerAsnPhelie-----TyRLeuIleLeuTyRAsnLysIleAsn 70

QY 246 AAAATCAAAAACAGATTCATTCATGATTCGACATTAAC---TGTAAATACATATCA 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 IleTyRAsnAsnLysTyRGlulIleLysAsnAsnTyRLeAsnValPheLeuAsn 90

QY 303 TTTTTCCTTAAGACACACACGCTTATTTGATTAATACACGCGAAATATATAT-- 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 nTyRtyTyRLeuLysValIleAsnLysIleGlnGlyIleLeuAsnAsnLeuTyRAs 110

QY 361 -----TTAAATCTGCAGACAAAAT-----GAACACACATTAATAATGAA 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 nLysIleAsnProIleTyRSeRAsnLeuPheLeuPheAsnAsnLysIleLysIle 130

QY 402 AGAGGCTACTAT-----TTAGCCAAAATCAAAATATGT 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 sTyRSeRGlLeuGlnGlnLeuIleGlyTyRlySglTyRlySeRAsnIleLysLys 150

QY 435 AATTTAGAAAACGTTTAAATCTTCATCAATTAATTAATTAACATGTTT----- 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 tIleTyRGlulSproValIleAsnAsnTyRLeAsnGlnLeuAsnIleTyRGlulTyR 170

QY 487 -----TTAAAGTGTTTAAATTTAATATATATGATTAATTAAGAAACCATATC 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 eleuSerCysTyRlySerLysLysLysIleIleAspThrAlaLeuLysThr----- 187

QY 540 AATTCACACATTAAGTAGAATAATTAAGTGTGTTATTTTAACAGAAATCAACTTA 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 -----AlaAspSerGlyTyRLeuThrLysArgLeuIleAs 199

QY 600 AACAGATTTAATATACGATTTTGTATAAGATATGATTAAGATAGATGAGACCTA 659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 nIleThrSerAsnPhelIleLysLysIle----- 208

QY 660 TGACTATGCTGTGAATGTTTGCAAAACCTCAACCTGACAACTTACATTCATTA 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 -----LeuAsnCys-----LysSerProPheIleLeuLysTyRleLeuAs 222

QY 720 TGACGATTAATATAGGATGATGCGCATATTCATTTATAGTATATTTTCCAAAT-- 774
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 n-----MetaPilietyRGlulysnIleI 230

QY 775 -TTTGACTATACAAATTAACATTTTCATGATATATGAAATCTTACCTTAATAATGA 833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 eleuProLeuAsnIleLeuArgPheLysIle---LeuGlnAsnAsnIleLeuAsn-LeuA 249

QY 834 AATGACACGTTTCAGTTACTCAACCAATATCAAGATATTAATAATATCTTATACATTA 893
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 snAsnGlyThrPheIleTyRThrLysAsnThrTyRlySerTyRleLeuAsnLysL 269

QY 894 TCATTTCCATGAT-----ATGAATTTTATGACATATTAATACATCTTCCTGGG 946
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 euLeuAsnLeuTyRAsnArgAsnIleTyRLeuAsnIleLysSerValTyRleu----- 287

QY 947 GTTTAAATATGATTAACACCAATATACATAAACAACTAATGATGAGCCTGTTTTCATA 1006
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 288 --CysAsnIleTyRAsnAsn-----IleCysAsnThrCys-----L 299

QY 1007 TTGACATCAATTCGAGTTATTCCTTAT-----GTGATGATCATAG 1045
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 euAsnTyRlySlnLeuTyRlyTyRAsnLeuGlyGlnHisIleGlyValIleSerSer 319

QY 1046 AAAAAATTCACATGTTTATCTTTTACGAACTACTACAGAACCAACG-----T 1096
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 lValIle-----SerGlnProSerThrGlnMetV 329

QY 1097 TATCCCTACTTTTATAGATGATGACAAATTTTTCATTTATATATGATGATAAGATG 1156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AlleuArgThrPhe-----HisAlaSerSerIleLeuLys---AspLysPheA 344

QY 1157 TATTTACGATGATTTATTAATTAATTAATTAATCAACGCTGATTTACGT---CAAATGAT 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 snPheAsnLysTyRleuIleTyRlySlyIleTyRleuTyRLeuAsnIleAsnLysIleP 364

QY 1214 TAAATATCATATATATGATTAATGATTCACGTTAATATCAAAATACATTAAGATGA 1273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 heLysLeuIleIleAsnPhelLysTyRlyIleAsnIleLysPheAsnLeuIlePheLeu 384

QY 1274 TTCAGACATTCAGCGGTATTCATTCATGACATATACGTTAATTCGTTGTTATATG 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 eLAsnLysIleLeu-----TyRAsnTyRAsnAsn---IleLeuPheG 397

QY 1334 AATGCAATCTTTCATGACGAGTATATTTTTCAAAATATTTTATTAACA----- 1389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 lutyRlySlyTyR-----IleLeuGlnAsnGlnTyRlyIleLysCysAsn 411

QY 1390 -----CAAGCTAAGTTAAACAAAC---AAATCAATATGACATTCACCTTACAGCATATC 1438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 heIleTyRAsnSerIleSerLysAsnPhelLysTyRAsnLeuAsnIleIleIleLys 431

QY 1439 ACATTAAGTATGATATCAACGACACACCACTCAATCAATGAGAGGTTATGTTA----- 1491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 yRLeuAsnAsnValIleLysTyRtyRAsnTyRSeRAsnIleGlnLeuIleLysAsnI 451

QY 1492 -----TCTAAAGTGTTTTAAATGATATATGATGATACCTGATTCGTCATCAT-- 1542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 leHisAsnLysTrpIleLeuTyRAsnIleTyRThrTyRtyRleuTyRtyRtyRHisIle 471

QY 1543 -----TTTAACTTATTCGCT-----TTAGATGATTAACAAATGACATATACATA 1585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 ySPheTyRAsnLeuTyRAsnLysGlyIleIleLeuAsnAsnAsnAsnLysTyRAsnV 491

QY 1586 TCATTAAC-----GGTTACAAAACACTGACGATATATATTTCTCTCATTTG 1636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 AlleTyRPhelLeuIleAsnTyRphAsnLeuPheSerAsnTyRtyRtyRlySlyTyR 511

QY 1637 TCACATTCAGTTTCATTTGTTATCTTATTCCTTCCATTACTTA----- 1683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 snAsnAsnTyRAsnPhelIleAsnSerAsnTyRtyRphelySlyMetAsnPhelIleLeu 531

QY 1684 -----ACGGAAGTGAATTTGACGACAAATTTATTTATGCGATACGATAGTTGT 1735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ySAsnPhAsnAsnIleGlnIleLeuAsnLysLeuPheTyRValAsnAsnIlePheIle 551

QY 1736 ATATGAATVCCGTGTGTTAAACCTTATGAAACCCAGCTTATTTGACCCGATACCTTAG 1795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 yTyRtyRtyRGlulSlySlySlyPheIleTyRleuAsnIleIleAsnAsnIleIleIle 571

QY 1796 GTAAATGG---GATATGAAAACGACAGATAGATTAAGATGTTTGACTG-----A 1843
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 yLysTyRleuAsnPhelTyRlySlyTyRtyRtyRAsnLysLeuPheIleLysLysTyR 591

QY 1844 ATCTTAAGAAATATGACATATGAAGT-----A 1870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 snAsnPhelLeuTyRleuTyRglulIlePheLysTyRAsnTyRtyRtyRtyRleuLeuA 611

QY 1871 ATGGAAG-----ATTAAATTCGCTTCGCTGATATCCGAAAACGCTTGGATACAA 1924
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 611 snAsnLysTyRAsnLeuTyRleIleTyRAsnAsnTyRlyIleLysTyRleuTyRlySlyR 631

```


QY	64	ATGGAACATATACCGTACATTAAGTTAAAGTGAACGCAAAAAACCA-----ACCAAA	111
Db	228	LeuUUleuLeuSerPheLeuLeuLeileAsnGluLysAsnGluLysIlePheGluGlu	247
QY	115	TATATAAACGTTACTATTCTGTAGCAATTTGGTGTTTAATGCTATGAATTT-----	168
Db	248	TyrLysSerIle-----SerIleTyrPheLysPhePheLysAsnPheLysIleLeuPhe	265
QY	169	-----GATGTGAAGTATTTCGAGTTTCGAACTCTTT	201
Db	266	IleLeuIlePheLysIleIleLysLysAspValPheLeuLeuLeuLysTyrLeuLysLeu	285
QY	202	TATGACGCAATTTTATACGTATGTGAAGAAGCGTGAATCAATCAAAATCAAAACAGAT	261
Db	286	LeuLysLeuThrTyrPheTyrIleLys-----SerIleLeuGluLysLeuAsn	302
QY	262	ATTATCATGATTTGGACATTAACGTGAATTAATACGATTAATCTTTTACTTAA-----	315
Db	303	LeuIleLeuTyr-----AsnIleTyrAsnAsnAsnTyrIleAsnLysIleAsnIle	318
QY	316	-----GACACCATGCGTTATTGTTGATTAATTTACACGC-----GAA	351
Db	319	TyrIleLeuThrTyrAsnThrLysLysLysTyrAsnAsnLeuIleLysTyrLysAsnAsn	338
QY	352	AATATATATTATTAATCTGCAAGAAAGAAATGAACACACATTAATAAATGAAA--GAGCGT	408
Db	339	AsnIleTyrLeuTyrAsnAsnAsnTyrSerLysTyrAsnAsnLysLysLysIleAsnIle	358
QY	409	ACTATTTTACGCAAAAT-----	426
Db	359	SerLysIleIleLysAsnPheLysIleAsnIleAsnTyrAsnAsnIleTyrIleTyrLys	378
QY	427	-----CAAAATGAATTTAGAA--	444
Db	379	LeuLysAsnLeuTyrAsnTyrIleIleAsnIleAsnThrLysGluIleIleLeuGluLeu	398
QY	445	-----AAACGTGTAATCTTCATCAATTTA	471
Db	399	LysIleLeuTyrIleAsnSerLysArgProPheLysGluLeuLysTyrAspLysSerLeu	418
QY	472	GATTTAACAATGTTTAAATGGTTTAAATTAATTAATTTATTCATCACTTATGAAA--	528
Db	419	TyrIleIleLysIleLeuAspGluLeuGluTyrAlaMetLeuThrAsnPheLysSerTyr	438
QY	529	ACCAATACATCAATTGCACATTAGGTACAAATTAATCTGATGCTGTTTAAACAGAA	588
Db	439	LysAsnLysAsnIleIleLysTyrHisLysLysLysTyrLysPheLysIle---Pro	457
QY	589	TCAACAACCTTAACAACATTTTAATTAACAGATT-----TTTGATAAGAATATGATAG	642
Db	458	SerLysIleGluLysTyrIleAsn---ThrIleLysProIleIleLysLysAsnIleIle	476
QY	643	AATGATAGTGAACCTATGACTATAGCTGCTGCAAAAGTTTGCAAAATC-----	690
Db	477	GluIleLeuGluLysLysLysIleLysIleAsnAsnTyrLysGluValIleSerLeuIle	496
QY	691	-----ACACCTGGAACAACCTTACATCTATCATATACAGCG-----ATTATATTA	735
Db	497	AsnIleAsnLysLysAsnLysIleThrTyrIleAsnLeuAspIleTyrSerLeuIleLys	516
QY	736	GGTATGGCCATTCATTTATAGTGAATATATTTCCAAATTTTGACTATACAAATTAACA	795
Db	517	IleAsnLeuTyrIleIlePheLysAsnIleIleProSer-----	529
QY	796	TTTTCATGAAATATATGAACTCTTACTGTGAATATCAAAATGACACGTTTCAGTTACTC	855
Db	530	---HisLeuAsnIleIleLeuIleLysLeuAsnAsnProMetLeuLeuIleLysSerIle	548
QY	856	AACCAATATCAAGATATTAATATCT-----TATACACATTTATCTTTCCATGATAGT	909
Db	549	---GlnTyrTyrAsnIleIleIleAsnLysLeuIleLeuAsnGluHisIleTyrIleIle	567

QY	910	AATTTTAT-----GACATAATTAAACATCTCATCGTGGGTTAAATATGTAAAC	963
Db	568	AsnIleIyrLysIleasnLeuIleLysaspIle-----LysIleLysLysTySer	564
QY	964	ACCMAATACATTAACAACAACATATGCATAGCCCTGTTTTTCATGCATCAATTCGAGT	1023
Db	585	LeuasnIyrllePhelysuIule-----LysIleGlnThpAsnLysIle	599
QY	1024	TATCCTTATGTGATGATATCATGTAAAAAATTCACAATGTTATACTTTACGAAACACTAT	1083
Db	600	ThrLysTyrlleasnIleasnGluLysIleIethrLysLeuIleLeuTrp-----	616
QY	1084	TCGAACCCAAGTTAAATCCCTACTTTTATAGATGCATCAATTAATTTTCATATATAG	1143
Db	617	-----ProIleuIleLeuasnPheIleLeuSeraspIleCysLeuSerIleSerPhe	634
QY	1144	ATTGATAAGAATGTATTTAATACGATGATATTATTAATTAATTAATCAACGTGTATACGT	1203
Db	635	LeuLysLysLys-----AsnLysLysLysIleLysLysIleasnSerLysIleasnLys	652
QY	1204	CAATGATTT-----GTAANAATCTATAATAATAGANAAT	1236
Db	653	AsnLeuLeuLysLeuasnLeuLysPheLysLysTyrlleLysTyrlle-----	668
QY	1237	GATTACGTTATATCAATACAAATACATTAAGAATGATCAAGACATACCGATTATGAT	1296
Db	669	---LyrIleasnVal-----IleLysLeuLeuLysLeuIlePhe	680
QY	1297	TGCATGCATATACGTGTTAAATTCGCTTTTATATATGAATGTAATACCTTCATGACAGCT	1356
Db	681	CysLysThrIleTyrlleIethrPheIleValAlaTyrlProLeuasnTyrlPheIleasn	700
QY	1357	GAT-----ATTATTTT-----CAAACTATTTATTAATAACA	1389
Db	701	GluPheasnLysTyrlTyrlIleIlePheLeuLysasnIleGlnAsnTyrlPheIle	718
QY	1390	CAAGGTAACTTAAAAAACAATAATCAATATGACATCACCTTACGACTATACACTACTGAT	1449
Db	719	-----LeuasnTyrlleuIleIe--LeuasnTyrlleuIleIe--	724
QY	1430	GATATCAACGACACCCATACCTCAATATGAGAGAGGTATGTATCAAGTGGTTTAAAT	1509
Db	725	-----AsnLysasnIleLeuLeuthrIleIethrIlePhe	736
QY	1510	GGATTA-----TATGCATACCCGCACTTACGTTGCAT-----	1542
Db	737	GlyLeuThrLeuThrIleLysTrpSerArgGluValHisLysIleTyrcylAsnTyrlLeu	756
QY	1543	-----TTTAACTATTCGCTTAGATGATGAATGAACAAATGACACTATACAAATTC	1587
Db	757	TrpThrMetThrIlePheProLeuPhe--LeuGlnSerHisLysIleLeuIyrylLeuLys	775
QY	1588	ATTACGGTTACAAAACACGATGAACGATATATATTTCTGCATACCTTTGTCACACTAGCT	1647
Db	776	LeuasnIleIleSerIleSerIleasnLeuasnTyrlleGluasnIleIleIethrLysLeu	795
QY	1648	TCATGTATTAATATTG-----GTTCCTTTCCAACTACTTAACGSAAGTGAATATAC	1701
Db	796	AsnLeuTrpAsnIleuLeuasnTyrlLys-IleGlnasnIleTyrrasnlIepheSerTyrlTh	815
QY	1702	GACATTTTATTTATTTGCGATACGATGATGTGTATATGAATACGCTGTTTAAACCTTA	1761
Db	815	IleIleIleIlePhe-----PheLeuIleTyrcylLysShisGluAlaPheArg-----	831
QY	1762	TTGAACCCCAGTTATTCGACCCGATACGCTTAGTAAATGGCATATTGAANAACGACAG	1821
Db	832	-----SerValTyrlalaen-----LeuLysAlaThrAr	841
QY	1822	ATAGATATAGATGTTTGTACATGAATCATAAAAATATGCATATGAAGTGATGGAAGAT	1861
Db	841	G-----TrrPlvsPheIle	845
QY	1882	AAATAATGGTTCGTGTGATACCGAAAMAAGCCTTGATACACAGCGTGCATTTTAAACC	1941

```

Db      845 elvstYrProleuLeuile-----lleLeuileglulleLeuyspr 859
QY      1942 TTGTACGTGACAAATTCCTTTGACGGTGCATTTATGAAAACAATTAAGTATCTATAAT 2001
Db      859 o---TyrLeuileglTyrilleSerValPheilleuTyrleuile----- 873
QY      2002 GAGCAAGGTCAATTCGATATATCCGCTCAAAACGTGAATTTGTATGCTAATGATAT 2061
Db      874 -----lleTyrTyrGlullelleLeuAsnleuAsnSerTyrillelleTyrAs 891
QY      2062 GATCAATATTTT-----CTGATGACCTTAATATGAAAGCGATTTTATTAAGACGCT 2118
Db      891 nileAsnillelleTyrleuLeuTyrille-----lleAsnGlTyrillelle 909
QY      2119 AGAGAAA-----ATTTTC 2130
Db      909 uProlYsileProTyrGlulYsleuileuValSerilleProAsnilleTyrleuileAs 929
QY      2131 GACCAATGATCAATTTGATGATTTCTTTATTTGAAGTGACATCGTTCATTTTCACTT 2190
Db      929 nillelleleAspillelleleuPheheilleSerGlySTrpleuile---leuTyrle 948
QY      2191 AACGCTTATTTCCAGCTGACGTTGACGTACATACATA 2224
Db      948 uAGTnTrTyrVallyleYsAspYslySTyrVal 959

RESULT 4
Q38545 PRELIMINARY: PRT: 575 AA.
AC Q38545:
DT 01-NOV-1996 (TREMBLEL 01, Created)
DT 01-NOV-1996 (TREMBLEL 01, Last sequence update)
DE 01-JUN-2002 (TREMBLEL 21, Last annotation update)
DE Bacteriophage phi29 temperature sensitive mutant T52(98) DNA
DE polymerase gene.
OS Bacteriophage phi-29.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC phi-29-like viruses.
OX NCBI_TaxID=10756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TS2(98);
RX MEDLINE=90370456; PubMed=2118623;
RA Blasco M.A., Blanco L., Pares E., Salas M., Bernad A.;
RA "Structural and functional analysis of temperature-sensitive mutants
RT of the phi 29 DNA polymerase."
RL Nucleic Acids Res. 18:4763-4770(1990).
CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMACTIC ACTIVITIES: DNA
CC SYNTHESIS (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT
CC DEGRADERS SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNDICLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE
CC + (DNA)(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL: X53370; CAA37450.1; -.
DR InterPro: IPR002064; DNA_pol_B.
DR InterPro: IPR004868; DNA_pol_B-2.
DR Pfam: PF03175; DNA_pol_B-2; 1.
DR PRINTS: PR00106; DNAPOLE.
DR SMART: SM00486; POLBc. 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN.1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ SEQUENCE 575 AA: 66827 MW; 8CADBFC7D5D50762 CAC64;

```

Alignment Scores:

Pred. No.:	5,46e-07	Length:	575
Score:	243.50	Matches:	151
Percent Similarity:	35.70%	Conservative:	115
Best Local Similarity:	20.27%	Mismatches:	238
Query Match:	5,98%	Indels:	241
DB:	9	Gaps:	40

```

US-09-727-892a-2 (1-2286) x Q38545 (1-575)
QY      31 AAACGTGAACGTGATGATGATTTTATCTGCGATATAGAACATTAAGCTACAAATAAGTT 90
Db      2 LysHismetProArGlysmetYrSerCysaspPheGluThr----- 15
QY      91 AACGACGCAAAAAAACCAACCAATATATAAAACGTTACTTATTCCTGATGCAATTGGTTGG 150
Db      16 -----ThrThrYsValGluAspCysArgVal-----Trp 25
QY      151 TTTATGCTTATGAATGATGATGATGATTTTCCGATTTTCGAA-----TCTTTTAT 204
Db      26 AlaTyrGlyTyr---MetAsnilleGluAspHisSerGlyTyrYsIleGlyAsnSerleu 44
QY      205 GAGCATTTTTATACGTATGTAAGAAAGACGTGATACATCAACAAATCAAAAAACGATATT 264
Db      45 AspGluPheMetAlaTrpVal-----leuYsValGlnAlaAspLeu 58
QY      265 ATCATGATGTCACATCACTGATGATTAATATAGATATATCATTTTACTTAAGAACACCATG 324
Db      59 TyrPhe-----HisAsnleu---LysPheAspGlyAlaPheille----- 71
QY      325 CGTTATTTGATATATATATACAGCGAAATATATATTTAAATCGAGAGAAATGAA 384
Db      72 -----AsnTrpLeuGluArgAsnGlyPheLysTrpSerAlaAspGly----- 85
QY      385 CACACATTAATAAAGCAAGAGCTACTATTTCGCCAAAAATCAAAATGTAATTGTA 444
Db      86 -----LeuProAsnThrTyrAsnThrIlele--- 94
QY      445 AAACGTGTAATCTTCAATCAATTTTATTAACAATGTTTAAAGTGTAAATTT 504
Db      95 SerArgmetGlyGlnTrpTyrMetIleAspIleCysleuGlyTyrYsGlyLysArg--- 113
QY      505 AATATATATGTAACCTTTATTAACCAACATATCAATGCAACATTAAGTGAATTA 564
Db      114 -----LysIleHisThrValIleTyrAspSerleuYsleu 126
QY      565 CTGATGCTGCTATTATTAACAGAACTCAACTTAACAGATTTTATTAATGATTTT 624
Db      127 -----ProPheProValLysIleAsnIleAlaLysAspPheLysleuThrValleu 142
QY      625 GATAAAGTATATGATATGATATGATATGATGAGCCATGACTACTGCTGGAATGTTTGA 684
Db      143 LysGlyAspIleAspTyrHisLysGluArgProValGlyTyr----- 156
QY      685 AAACGACACCTGAACCACTTACATCAATCATATGATGACG---ATTATTTAGTATG 741
Db      157 LysIleThrProGluGluTyrAlaTyrIleLysAsnAspIleGlnIleAlaGluArg 176
QY      742 TGCCATATTCATTTATAGATATATTTCCAAATTTGACTATACAAATTAACATTTTCA 801
Db      177 LeuLeuileGlnPheLysGln-----GlyLeuAspArgMetThrAlaGly 191
QY      802 TTGAATTTTANGAATCTTACTTGAATATGAAATGACAGT-----TTT 846
Db      192 SerAspSerleuYsGlyPheLysAspIlelleThrThrYsLysPheLysValPhe 211
QY      847 CAGTTACTCAACCAATATCAAGATATTAATAATCTTATCAACATTAATTCATTTCCATGAT 906
Db      212 ProThrleuSerleuGlyLeuAspLysGlnValArgYrAla----- 225
QY      907 ATGAAATTTTATGCTATATTAATCATTCATGCGTGGTGTAAATATGTAATACACC 966
Db      226 -----TyrArgGlyLysPheThrTrpLeuAsnAsp 235
QY      967 AAATACATTAACAACATTAATGATGAGCTTGTTTTCTATGACATCAATTCGAGTAT 1026
Db      236 ArgPheLysGluYsGluIleGlyGlnGly---MetValPheAspValAsnSerleuTyr 254
QY      1027 CCTTATGTGATGATATGAAAAAATTCACACATGTTATACTTTTACGAACACTATTC 1086
Db      255 ProAlaGlnImetYrSerArgleuLeuPro-----TyrGly 266

```


QY	1087	GAACCAACGTTAAATCCCT---ACTTTTATGATGATGACAATTAAT-----	1128
		::: :::: :	
Db	267	GIUPRollevalPhegiuglyLysTyrrValITPaspGluasptyrProLeuHisIleGln	286
QY	1129	-----TTTCATATATPAGATTGATGAAGATGATATTAAGCATGATATA	1173
Db	287	HisIleargcysGluPhegiuleu-----LysGluGlyTyrIleProthrIle	302
QY	1174	TTAATTAATAATTAATACGTGTATTAACGTCAATGATTTGTAAATCTATATATATGAT	1233
		:::: :	
Db	303	GlnIleLys---ArgSerArg-----PheTyrLysGly	312
QY	1234	AATATTATACGTTATATATACATAACATTAACATTAAGATGATTCAGACATTCGGGGTAT	1293
		: ::: :::: :::: :	
Db	313	AsnIuIyrlleuLysSerSerIyrgLysGluIleIlealAspLeuTyr---LeuSerIasVal	331
QY	1294	GATTGCATGCATATACGTGTAAATTCGTTGTTTATATATGATATGCATATAC-----	1344
		::: :::: :	
Db	332	Asp---LeuGluLeuMetLysGluHisTyrAspLeuTyrAsnValGluIyrrIleSerGly	350
QY	1345	-----TTTCATGCACGTGTATATATTTATTTTCAAACTAT-----TTTAT	1383
		: : :::: :	
Db	351	LeuLysPheLysIalThrIthIrgLysLeuPheLysAspPheIleAspLysTyrPheThrIle	370
QY	1384	AAAAA-----CAGGTAAAGTTAAAAACAAATCAATATGACATCACCTTAGCACTAT	1437
		: :	
Db	371	LysThrIthrSerGluGlyIalIleLys-----	379
QY	1438	CACATTACTGATGATATATCAACGAACCCACTACTCAATAGAGAGGTATGTTACTATAA	1497
		----- :	
Db	380	-----GlnLeuIalLys	383
QY	1498	GTCGTTTAAATGATATATATGCG-----ATA	1524
		:	
Db	384	LeuMetLeuAsnSerLeuTyrGlyLysPheIalSerAsnProAspValThrGlyLysVal	403
QY	1525	CCTGCATATACGTTACATTTTAACCTTA--TTCCGTTTATGATGATATACATGAACTATAC	1581
		: : :	
Db	404	ProTyrLeuLysGluAsnGlyIalIleGluLysPheArgLeu-----	416
QY	1582	AATATCATACGGTTTCAAAAACACGACGAACGTAAATATATA-----TTCTCTACATTT	1638
		: :	
Db	417	-----GlyGluGluGluIthLysAspProValIyrrThrProMetGlyIalPhe	432
QY	1636	GTCACATCACGTCGTTATGTTAACTTAATTTGGTTCCTTCCCATACTTAAACGGAAGTGA	1695
		: : :	
Db	433	IleThrIalThrIalAspIalAspGlyThrIthIleThrIalAlaGln-----Ala	447
QY	1696	ATTGACGACCAATTTTATTTATTTGCGATACGATAGATTGATATGGAATCCGTTGTAA	1755
		: :	
Db	448	CysTyrAspArgIleIleTyrCysAspThrAspSerIleHisLeuThrGlyIthrcIule	467
QY	1756	CCCTTATTTGAACCCACAGTTATTTGACCCGATAGCCTTAGTAATGGATATGAAGAC	1815
		: : :	
Db	468	ProAspValIleIleLysAspIleValAspProLysLysLeuGlyIyrrThrIalHisGluSer	487
QY	1816	GAAACGATACATTAAGATGTTTGATCTGAATCATTAAGAAATATGACATATAGAGGATGA	1875
		: :	
Db	488	---ThrPheLysArgValLysTyrLeuIalGlnHisTyrIrrIleGlnAspIle-----	504
QY	1876	AAGATTAAATTCGTTCTGCTGTATACCGAAAAACCCCTTGATACAAAGCGTCATTTT	1935

Db	504	-----	504
QY	1936	GAACCTCTTGACGAGGAACAATTTGTTGACGGGCCATTTATGAAGAACATAAAATATAC	1995
		: : :	
Db	505	-----TyrMetLysGlu-----ValAspGlyLysLeuValIuIuLysSerProAspAsp	520
QY	1996	TATTAATGAGCAAGGTACATATGCATATATCCGCTTAAAAAGTAATGTATGTGCGTAAT	2055
		: : :	
Db	521	TyrIthrAsp-----IleLysPheSerValLysCysAlaGly	532

QY	2056	GTATATGATGATATTTTACTGATGAGCACTTAATATGCAAGCTCAATTTATTAATAAGC	2116	1121
Db	533	MethrapsDys-----	2116	1121
QY	2116	GCTAGAGAAATTC	2130	1121
Db	546	phelysValglypne	550	1121
RESULT 5						
QY	096192	PREDIMINARY;	PRT:	1121	AA.	
AC	096192:					
DT	01-MAY-1999	(TReMBLrel, 10, Created)				
DT	01-MAY-1999	(TReMBLrel, 10, Last sequence update)				
DT	01-DEC-2001	(TReMBLrel, 19, Last annotation update)				
DE	Hypothetical 135.8 kDa protein.					
GN	PF80495W.					
OS	Plasmodium falciparum.					
OC	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
OX	NCBI_TaxID=58333;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=99021743; PubMed=9804551;					
RA	Gardner M.J., Tettein H., Garucci D.J., Cummings L.M., Aravind L.,					
RA	Shoen K.E., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,					
RA	Keen C., Jing J., Aston C., Lai Z., Schwartz D.C., Petrea M.,					
RA	Falberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,					
RA	Frazer C.M., Adams M.D., Venter J.C., Hoffman S.L.					
RT	"Chromosome 2 sequence of the human malaria parasite Plasmodium					
RT	falciparum."					
RL	Science 282:1126-1132(1998).					
DR	EMBL, AF001398; AAC71888.1; -					
KW	Hypothetical protein.					
SQ	SEQUENCE 1121 AA; 135780 MW; 6ABAE17AB33D415B CRC64;					
Alignment Scores:						
	Pred. No.:	6.4e-07	Length:	1121		
	Score:	242.00	Matches:	201		
	Percent Similarity:	37.99%	Conservative:	136		
	Best Local Similarity:	22.66%	Mismatches:	291		
	Query Match:	5.95%	Indels:	261		
	DB:	5	Gaps:	52		
US-09-727-8924-2 (1-2286) x 096192 (1-1121)						
QY	100	AAAAAACCAACCAATATTAATAAACCTTACTTATCTGATGCAATTGGTGGTTAATGCT	159	1121
Db	73	LysAsnLeuPcIyStYrLysCysAlaLysTyRglucYsIleSerAla-----	88	1121
QY	160	TATGAAATTTGATGTTGAAGTATTTCCGAGTTTCGAATCTTTATGAGCATTTATACG	219	1121
Db	89	-----LysGluValfYrLysTyRLeuAspGluYrYrLysCysPheAsn	104	1121
QY	220	TAT-----	252	1121
Db	105	YrIleSerLeuCysAspIleIleGlnSerYalIlePheAspGluAspLysfYr	124	1121
QY	253	AAAAACAGATATTATCATGATGACATACATGATATAATTCGATTAATCATTTTACTT	312	1121
Db	125	PheThrAspTYrAsnPhetYrIleGluValLysAsn--IleAspLysValLeuAsn	143	1121
QY	313	AAAGACACATCCGTTATTTGATTAATAT-----	359	1121
Db	144	LysIleAsnGluIleTyR-PheLysAsnLysAspIleThrPheThrAspArgGluIleLeu	163	1121
QY	360	TTTAAATCTGCAGAGAAATGACACACATTAATAATGAAAGAGGC-----	407	1121
Db	164	GlyLysIleCysAsnLysIleMetSerYrIleHsGluMetAsnGlyAsnGluIleLeu	183	1121
QY	408	-----TACTATTTTAGCCAAATACAAATGTAATTTTAAATAAACGTTTAATAC	458	1121
Db	184	HisPheLeuIleTyR-PheThrAspGlyTrpAsnAsp-----LysAsnLeuIleLeu	201	1121

QY	459	TTCAATCAATTAGATTATTAACAAGTGTTTTAAATGGT--TTTAAATTAT-----	507
Db	202	PhetyrAsnTyTyrPheAsnTyValPhe-ASPHISMeTyrLeuPheAsnHisGluI	221
QY	508	-----ATTATGATACTTATGAACCAATCAATCAATTCGCAACAT	551
Db	221	eTyTyrLeuLeuPheIlePheAsnLysTyTyrLeuAsnAspSerAsnIle---ProPh	240
QY	552	ACGTAGAATTTACTGAGGT-----GCTTATTTAAACAAGATCAACACT	596
Db	240	eAsnLysAsnLeuIleGlnGlnMetGluPheAsnLeuTyTyrPheArgGlu-----	258
QY	597	TAAACACATTTTATTTATACAGTTTGTGTTAAAGATATGATATGATGTAAAC	656
Db	258	eLysAsnGluLysAsnTyTyrIleIle-----LysMeAsnLysGlyGluI	273
QY	657	CTATGACTATGCTGGAATGTTTTGCAAACTCACACTGGAACAACATACATCA	716
Db	273	eTyTyrLys-----LysCysPheAlaLysPhe-HisGluAsnValAspHisIleAs	289
QY	717	TAAAGACGTGATTATA-----	732
Db	289	PasnGluLysIleLeuAsnIleLeuA-gLeuTyTyrValAspAsnSerIleLeuAspIleAs	309
QY	733	-----TTAGATATGGCCATATTCATTTAGAGATATTTCCAAATTTGGACTA	782
Db	309	PleAsnAsnLysMetLeuCysAsnLeuAsnAsnLeuIleAsnGluAsnIleGluTy	329
QY	783	T--AACAAATTAACA-----TTTTCATTGAATATTATGCAATCTTACTGAAATATGA	833
Db	329	IleSerLysLeuLeuAsnAsnPheTyTyrCstIleLeuIleLysGlyLysTyTyrAspAsnAs	349
QY	834	AATGACAGCTTTTCAGTTACGTACCAACCAATTCAAAGATTAATATCTGTATACACATTA	893
Db	349	PMeThrIleLysIleLysLeuLysGlu-----ValIleLysAlaIleThrHisH	364
QY	894	TCATTTCCATCAT-----ATGAATTTTATGACTATATTAATCAATTC--TATCGTGG	944
			:::
Db	364	SileLeuCysAspLysThrLysAsnLeuGlnIuThrPheCysSerAspIleAspTySerTh	384
QY	945	TGGTTTAATATGTATTAACCAACCAATTAACA---AACCAACATATGATGACACC--TT	997
Db	384	LeuLeuAsnSerLeuAsnAsnLysPheIleLeuAsnLysIleIleAspLysPheIle	404
QY	998	GTATTTCTATTGACATCAATTCAGATTA-----	1027
Db	404	eLeuPheTyTyrGluCysLeuLeuLysIleLeuAsnIleLysPheValAsnPhGlnSe	424
QY	1028	CTATGTGGAT-----GTATCATGAAGAAAATTCACACATGTTATATCTTTACGACA	1081
Db	424	LeuCysLysSerLeuLeuIleSerLeuLysAsnIleTyTyrAsnIleLeuArgAsnAsnVa	444
QY	1082	ATTGAGAACAACCACTTATATCCCTACTTTTATAGATATGACATATTTTGTATATAT	1140
Db	444	Ityr-----IleValAsnAsnValLeuPhe-AsnAspIleMetLysPheSerLeuTyL	462
QY	1141	-----AAGATGATATAAGATGATATTATTAACGATGATTTATTA	1177
Db	462	eUcysAsnIlePheLeuGlnGlyLysArgIleLysThrGluAsnGluAsnAlaValLeuIleI	482
QY	1178	TT-----AAATTAATACACCTGATATAC	1201
Db	482	LeHisAsnAsnAspGlnThrAsnTyTyrSerAsnLysGluAsnIleLysAspIleIleG	502
QY	1202	GTCAAAATGATTTGTTAAATAC-----TATATAAT-----	1230
Db	502	IuLysArgIleLysGluTyTyrIlePheTyTyrLysMetGluAsnTyTyrLysAspPheHisPheL	522
QY	1231	-----GATATGATTTAGCTATATATCAAT-----ACAAATTCATTAAGAATGATCT	1276
Db	522	ysLeuLysAspSerAspLeuLeuSerIleLysLeuLeuSerAsnThrPheValLysIleA	542
QY	1277	AAGACATT-----ACGGGATGTGATTCATG-----	1302

[illegible]

D	518	AspProGlnTyrAsnLysTyrMetPheAsnAsnGluSer-----Phe	531
Q	1504	TTAAATGATATATATGGCATACCTGCATACGTTCACATTTTAACTTATTCGGTTAGAT	1566
D	532	IlleLysAsnIleAsnArgLysArgProIleIleHisAsnTyrAspTyrGlnTyrPleuAsp	551
Q	1564	GAT-----AACATGAACATATACAAATTCATTACAGCGTTACCAAAACACTGACGT	1614
D	552	AsnSerLeuLeuAsnHisLeuIleLysAsnIlePheLysGlyTyrLys-----	567
Q	1615	AATATATATATCTCTACATATTGGCACATGCACGCTCATATTGTATAACTTATGTCTCTTC	1674
D	568	-----TyrSerLysTyrIleIleLeuAsnLysLeuTyrSer-----	579
Q	1675	CAATACTTAAACGGAAGATGAATTCAGCACAATTTATTTATTCGATACGATAGTTG	1734
D	580	AsnTyrLeuPheAsnSerThr-----IleTyrCysAspAsn-----	591
Q	1735	TATATGAATTCGGTGTGTTAAACCCCTATATGAAACCCAGTTTATTCGACCCGATAGCTTA	1794
D	592	-----LysIleIleGlnAspIleLysIleAsnSerThrLeuTyr-----	604
Q	1795	GGTAAATGGGATATATGAAACAGAACAGATAGATAAAGTATTGTGCTAGATCATAGAAA	1854
D	605	-----LysTyrIleCysTyrGlnAsnLys-----	612
Q	1855	TATGCATATAGACGATGAAGTGAAGATTAA-----	1884
D	613	AsnCysLeuAsnValAsnSerLysIleSerAsnGlnAsnAsnIleAsnIleLysAsn	632
Q	1885	-----ATTGCTTCGCTGGTATACCGAAAC-----	1911
D	633	AsnLeuCysIleTyrGlnGluProThrValProLeuLeuAsnLeuProAspAsnIleSer	652
Q	1912	-----GCCTTGATACA-----ACGCTGATTTTGAAACCTTT	1944
D	653	LysLeuIlePheAspLeuAsnIleGlnAsnIleIleTyrAsnIleAspLeuSerAsnIle	672
Q	1945	GTACGTGAACAATTC-----TTTGACGGTGGCCATTATGAA	1980
D	673	AsnIleAsnGlnTyrIleAspIleTyrAsnAsnValLeuPheAspIleValIleLysTyr	692
Q	1981	AACATATAAAGTATCTATATATAGCAAGGTACAAATGATATACCGCTGTAAACGTGA	2040
D	693	AsnAsnIleAsnLeuTyrAsnTyr-----IleIleLysLeuTyrPro-----	706
Q	2041	ATTGTATGTGGTATGATATATATGATGATATTT-----ACTGATGAA	2082
D	707	-----TyrTyrAspLysTyrPheIleLysAspIleAsnThrProTyr	721
Q	2083	CTTATATATGAACGTGAATTTATATATTAAGACGGCTAGAGAAATTCAGACATAGTCA	2142
D	722	MetCysLysTyrIleLeuIlePheTyrAsnAsnTyrThrThrIleAsnIleIleAsnAs	741
Q	2143	TTTGATGCAATCTTATATATTTGAAGTGACATCGGTTCATTTCATTTACATTACGACTTATTT	2202
D	741	AsnAsnIleSerAsnIleLeuSerAspAsnLysIleGlnTyrSer---ThrIleValTyr	760
Q	2203	CCAATGACAGCTCAGTACATACATACAAATGCGTT	2236
D	760	rgluIleAsnAsnThrIleValSerAsnIleIle	771
RESULT 7			
ID	096204	PRELIMINARY; PRT; 3973 AA.	
AC	096204;		
DT	01-MAY-1999 (TREMBLrel. 10. Created)		
DT	01-MAY-1999 (TREMBLrel. 10. Last sequence update)		
DE	HYPOthetical 480.3 kDa protein.		
GN	Plasmodium falciparum.		

OC Eukaryota: Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5633;
RN
RX (11)
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettein H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petrea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.,
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum".
RL Science 282:1126-1132(1998).
DR EMBL, AF001402; AAC71900.1;
SQ Hypothetical protein.
KW
SEQUENCE 3973 AA; 480283 MW; 9A7ED9E1196213B5 CRC64;

Alignment Scores:
Pred. No.: 1.25e-06 Length: 3973
Score: 236.50 Matches: 190
Percent Similarity: 38.77% Conservative: 143
Best Local Similarity: 22.12% Mismatches: 353
Query Match: 5.81% Indels: 173
DB: 5 Gaps: 48

US-09-727-892A-2 (1-2286) x 096204 (1-3973)

OY 10 CTGACATGC-----ATGCATATCATTAACATGACGTGCAATGATTTATAC 57
Db 2163 LeuAsnCysLysAsnGlyTyrPheLeuAsnLysAsnLysAspPheLysGluLeuTyr 2187
OY 58 TGGGATATAGAAACATAGCGTACAT----- 84
Db 2183 Tyr-----LeuTyrAsnLysAsnLysLeuSerGluLeuLeuLeu 2197
OY 85 -----AAAGTTAAACGAGCAAAAAACCAACAATATATAAAGCTACTAT 132
Db 2198 IleLysAlaIleTyrIleAsnGluThrLysIleTyrProLeuIleLeuAsnIleCysTyr 2217
OY 133 TCTGTACGAATGCT-----TGGTTAATAGCTTGAATGATTCGTAATTTTCG 186
Db 2218 AspArgAsnIleSerAsnIlePhePheAsn-----IleSpryTrsAsnLeuAsn 2233
OY 187 AGTTTCAATCTTTTATGACGCATTTTATACSTATGTGAAGAAGCGTATCAATCA 246
Db 2235 SerIleLeuGluTyrThr-----TyrLeuAlnLysLysLysLysAspHisLeu 2252
OY 247 AAATCAAAAACAGATATTATCATGATTCGCATCAATC-----TGTATTAATAC 294
Db 2253 AsnLeuLys-----TyrLeuLeuCysLysAsnLysSerIleHisMetHisLysTyr 2265
OY 295 -----GATATCATTTTATTACTTAAAGACACACATGCGTTATTTGATAT 339
Db 2270 IleSerTyrIleAspAspAspHisLeuIle-----AsnAsnMetLeuHisLeu----- 2285
OY 340 ATTACACAGCAAAATATATATTAAATCTGCAGAGAAGAAAAGAACACATCAAAATG 399
Db 2286 LeuArgTrsLysAsnIleTyrTyrLys-----TyrValLeuAsnIle 2299
OY 400 AAAGAGCTACTATTATTAGCCAAAATCAAAATGTATTTTGAAAAGACGTGTTAAATCT 459
Db 2300 AsnGluTyrAsnAsnPheLeuAspAsnHisLysCys-----LysArgLysArgLys 2316
OY 460 TCAATCATTTAGAT-----TTAACAAAGTTTAAATGATGCTTTAA-----TTTATATTT 510
Db 2317 PheIleAsnLysTrsAsnAsnIleGlnIleSerTyrAsnAsnAsnTyrAsnIleTyrAsn 2336
OY 511 ATTATACCTTATGAAAACCAATACATCAATTC-----ACATTAGTAGAGAA 561
Db 2337 ThrAsnAsnPheTyrGluTyrHisAspTyrIleAlaIleLysAsnIleLeuHisLysLys 2356
OY 562 -----TTACTGTATGCTGCTGTTTAAAGATCAACAACCTAAAGCATTTTATATAT 615

RT flesh-eater."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL: AP003186: BAB80075.1: -
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 1301 AA, 154624 MW, F327A5A51610D1A3 CRC64;

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Query Match:	Gaps:
3,08e-06	1301	230.50	189	156	19.77%	276	335	5,66%	16

US-09-727-892a-2 (1-2286) x Q8XNG4 (1-1301)

```

QY 151 TTTATGCTATGAAATTTGATTTGAAGTATTTCCAGTTTGACATCTTT---TATGAC 207
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 PheasnlystYrleuasnpluilemetYrPhe-----PhegluileuYstYrAsp 301

QY 208 GCATTTTAT----- 216
    |||
Db 302 lIleValPhePhegluAspIeuaSPArGpHeaSPaSnleuGlulIePheThrIySleuArg 321

QY 217 -----ACGTATGGAAGAAGCCGTGATACATGACAAATCAAAACAGATATTATC 267
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 GluleuasnThrleuileasnlySAlaGluseRileSerArglySValThrPheValItyr 341

QY 268 ATGATTCGCACATTAAGTATTAATACGAAATCATCTTTTACTTAAAGACACATGCGT 327
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 342 AlAlIeLyS-----AspGlulIePhePheIleYsGlueSnleuGlul 355

QY 328 TATTTTATATATTACACCGCAAAATATATTATTTAAATCTGCACAGAAAATGACAC 387
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 356 AspGlueSP-----GluTyRlySTyRileaspIySglueMetasnlySasn 371

QY 388 ACATTAATAATGAAGAG-----GCTACTTATTTTACCAAAATTCAA 429
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 ArgThrIySPhePheaspPheIleIleProValIleProIleValasnlyglueasnSer 391

QY 430 AATGTAATTTTGAAGAAAGCTGTAAA----- 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 TyRgluleuSerlySlySileglueInPheasnlySlyTyRlyValIglueInlySer 411

QY 457 -----TCTTCATTCATTTAGATTAGATTAAACATGTTTAAATGTTTAAATTT--- 504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 lIleIleSerlySglueuSeraspIeuaSPeSerMetPheIleaspSPeMetArgleuLeu 431

QY 505 ---AATATTTATGATTAACCTTTATG----- 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 ThrAsnIleTyRasnGluePheleuIleTyTyRlySlySleuValIleglueArglySasn 451

QY 526 AAAACCAATTCATCA-----ATTGCAACATTTAGTGAAGAAATTACTTGATGCTGT 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 452 LysThrIeuaSerSeraspSnleuIleuAlaIleIleValItyRlySasnleuTyRProVal 471

QY 577 TATTTAACAGATACACAACTTAACAGATTTTAAATTAATACATTTTGAAGATAAT 636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 AspPheThrySleuGlueInasnArglyglueValItyAsnValPheSerGlueSasn 491

QY 637 GATATGAATGATGTAACCCATATGAC----- 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 AspIleAlaSPArGAlaValIhlySlySleuasnlySglulIeYsGlueSlyArgThryasn 511

QY 664 ---TATGCTGTGAATGTTTGGCAAAACTCACACCTGACCAACTTACATCATCTTAAT 720
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 512 lIeTyHlIeueInlySglulIeueInlySasnlyglueInlyglueIleuTyRleuTyRasn 531

QY 721 GACGTATATATTTAGGTATGTCCTATTATCATTTAGTGAATATTTCCAAATTTTGAC 780
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 AsnGlulTyRleuSerasnlySmetAsnPheIleArgThr-----ProasnlySThr 548

QY 781 TAT-----AACAAATTAACATTTTCATGAAATATATATGAAATCTTACTTGAATATGA 834
    |||

```

```

Db 549 TyRSeRlIleaspIySleuSer-----aspPheasnValIleglueInlySasnSerasn 567
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 ATGACACGTTTACGTACTCAACCAATAT-----CAAGATATTAA---ATATCTTAT 885
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 568 SerIleArgTyRSeRValSerasnSPTyTyRThryasnSPTyRlySasnIleSerPhe 587

QY 886 ACACATTTAT-----CATTTCCATGATATGATTTTATGACATATTAATCAATTC 936
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 GlueSnPheMetThryIleasnlySlySlySleuasnleuTyRasnArgleuIleAlaIle 607

QY 937 TATCGTGTGCTTAAATATG-----TATACCCCAATATACATAACAA 981
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 608 LysSerlySglYlySasnIleglueInlySlySlySleuValAlaThrleuMetlySasp 646

QY 982 CTAATTTGATGACCCCTGTTTCTATTTGACATCATCTTG----- 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 627 AlaIleglueInlySlySlygluePheIySasnSerValAlaThrleuMetlySasp 646

QY 1020 ----- 1020
    |||
Db 647 GlueArgPhePheaspCySleuasnpluYsIleYslySglueAlaIleIleArgleu 666

QY 1021 ---AGTTATCCTTATGTC-----ATGATATCTGAAAAATTCACACATGTTTACTTT 1071
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 667 leuTySTyRlyTyRileaspIueTyRlySerTyTyRleu-----leuTyPhe 683

QY 1072 TACGAACCTATTCAGAACCAACGTTAATCCCTACTTTTAAAGATGACATATATTT 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 684 TyRglu-----GlyArgleuThrInglueIySaspPhegluPhe 695

QY 1133 TCATTATATAGATTGATTAAGATGTATTTAAGATGTTTATTAATTAATTAATTAAT 1191
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 696 -----ValGlnSerValIleTyRasnlySPoleuasnCySgluePheIySleu 711

QY 1192 CCGTATTAACGCAATGATGTAATAATAC----- 1221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 712 GlueSnIleasnGlulIleuVallySPheArglyIleglueaspPheglueSerlySAlaIle 731

QY 1222 -----TATATATATGATTAATGATTTACGTTATATATCATATACAAAT 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 732 leuasnTyRaspIeuleIleYSPheIeuaSPeSerPheIleglueIleuCySlySasn 751

QY 1261 ACATTAAGA-----ATGATTCAGACAT----- 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 752 ArgleuIySlyIleValGlyIleuSerGlueIySleuIleglueaspIleasnPheIle 771

QY 1285 -----ACGGGTATGATTCGATCGATATACGT 1311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 772 MetaspPheIleTyRlySasnlyglueIleaspSerPheIleglueIleuCySlySasn 791

QY 1312 GTTATTCGCTT-----GTTATATATGATGATGATCTTCAATGCA----- 1353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 lIeasnasnPheTyRPaSnasnIleTyRglueSnSerTyRleuProSerGlueIySlySasp 811

QY 1354 -----CGTGATATATTTTCAAAAC--- 1374
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 812 TyRmetIeugIleuLeuPheIySTyCySerIleglueIleuIleuIleglueIleuIle 831

QY 1375 -----TATTTTATTAACACAAAGT--- 1395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 832 GlueglTyRlyIleYsglTyRlyIleglueIleuasnSerasnPheleuIySThrIlegllyle 851

QY 1396 -----AAGTTAAAAAC-----AAATCATATG---ACATCACT 1428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 852 LysIleglueIySThrIySleuIySaspIleuIleIySleuasnIleIySleuIySasp 871

QY 1429 TACGACTATTCATTTACTGATGATATACGACACACCATCTCAATAGAGAGCTTATG 1488
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 872 lIeaspIeuaSPIleValaspIleIySasnGln-----SerasnlySlySTyMet 889

QY 1489 TTATCTAAGTCGTTTAAAGATTTATATGCACTACCTGATACCTTCAATTTTAAAC 1548
    |||

```

```

Db 890 -----ServalpnegiuasplleTy----- 896
QY 1549 TTATTCGGTTAGATGATACATGAACATACATATACATATACATGTTACGTTACAAACACT 1608
Db 897 -----Tyrasnaspneutyraspleu----- 903
QY 1609 GAACGTATATATATCTCTACAC-----TTTGTCCATCCAGCTTCATTTGATATAC 1659
Db 904 -----AsnillealalevalserLysIlellepheleleuglnslnslnlelelgluasn 921
QY 1660 TTATTC-----GTTCCTTCATCACTTAACGGAAGTGAATTAACGACATTTT 1710
Db 922 IleeraspvalgluleuserTyrserPheleleerLysasnlglnuleasnnyrleulle 941
QY 1711 ATTATTCGATCTGAT---AGTTGTATATGAATCCGTTGTTAAACCTTATTTGAC 1767
Db 942 LysTyrilleaspnasnilleasnleutyrilleatrgasnile----- 955
QY 1768 CCCAGTTTATTCGACCCGATAGCCTTATGTAATGGAT-----ATTGAA 1812
Db 956 -----Pheilegluproillevalillecluasnthrciuhtrilleleasplleleuasn 973
QY 1813 AACGACAGATAGATAGATGTT-----GTACTGATCATTAAGAATATGATAT 1863
Db 974 Asnlglnluvalasplysaspleulleaspasnillevalasnasnlysgln-----Phe 991
QY 1864 GAATGATGGAAGATTAATATGCTTCGTCTGTACCGAATAAGCGCTTGAT--- 1920
Db 992 Lysvalaspsasplleasnlyllegluasnleasplletrpansnullepheasn 1011
QY 1921 -----GTCTCTTCCTCAATGATGATGAT-----ACAGCGTGCATTTT 1935
Db 1012 LeulysvalasnserseTyrpgluasnilleleuserTyrlystrhrasnasnSerille 1031
QY 1936 GAAACCTTTGACGTGAACATTTCTTGACGCTGCATTTATGAAAC----- 1983
Db 1032 Gluthrcysleu-----GlygllypheleleasnasnProgluvalcys 1045
QY 1984 -----AATPAAGTATCTATAT-----GAGACGATGATCA 2013
Db 1046 LysgluleuleuLysaspthleutyrLysgluleserlleasnasnaspglulysglulle 1065
QY 2014 ATATGATATATCCGCTAAACTGAATGTGATGTGATGATGATGATGATTTT 2073
Db 1066 Ileaspleuphevalglulys-----PhevalmetserLysilleleaserasplulle 1083
QY 2074 ACATGATGATTAATGAAAGCGTGAATTTATTAAGACGCTAGAGAAATTTGAC 2133
Db 1084 ValaspleuleuserCysLysleuglyllethrphethrnxp-----Pheasp 1099
QY 2134 CATAGTCAATTTGATGATATCTTATATGAAAGTGAACGATCGTTCAATTTTCACTTAC 2193
Db 1100 -----Phegluaspsmetleu-----AspserrargilleleuvalleuLysasnlys 1114
QY 2194 GACTTATTTCCAGTTGAACGTTTCAGTACATAC-----AAATCTGATTTGCAT----- 2241
Db 1115 AsnleullealaleuthrcluserilleatrgalaleuLysasnaspTyrHISglyleu 1134
QY 2242 -----AATTTAAACGTGAACATGATGAATAAAGGCAAC 2280
Db 1135 HistleasleulleleglulysasnilleaspInpheleleaserglyasn 1150

```

```

RESULT 9
096170
AC 096170; PRELIMINARY; PRT; 1802 AA.
DT 01-MAY-1999 (TRIMBLREL 10, Created)
DT 01-MAY-1999 (TRIMBLREL 10, Last sequence update)
DE 01-MAY-1999 (TRIMBLREL 10, Last annotation update)
DN Hypothetical 216.8 kDa protein.
GN PF80375M.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

```

OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteil H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AE001390; AAC71864.1; -.
KW Hypothetical protein 1802 aa; 216824 MW; BE2A8448A81957AE CRC64;
SQ
Alignment Scores:
Pred. No.: 3,22e-06 Length: 1802
Score: 230.00 Matches: 202
Percent Similarity: 33.77% Conservative: 131
Best Local Similarity: 20.49% Mismatches: 242
Query Match: 5.65% Indels: 411
DB: Gaps: 59
US-09-727-892A-2 (1-2286) x 096170 (1-1802)
QY 7 TTACTAGAAATGATGATATATCATATA-----CATGAA 39
Db 931 LeutyrLysCysLeuGlnTyrLeuSerLysLysAsnAspleuThrLeupheilleasnlgln 950
QY 40 CGTCGATGATTTTATCTGGGATATA---GAACATTAACGCTACATTAAGTTAAGCA 96
Db 951 ThrGlullelekettyrleuasnillevallyLysleuLysglulnrgylsilleasn 970
QY 97 CGAATAAAACCAACCAATTAATAAACGTT---ACTTATTCCTACCAATTTGTTGTTT 153
Db 971 Ileasnlgln-----LyspheLysasnilleProasnHISleuLysglintlysglulle 988
QY 154 AATGTTATGAAATGATGTTGATGATTTCCGATTTGCAATCTTTATGACGATTT 213
Db 989 LysGlnValLysLysLysleuLysleuLysglulysnThrLysasnilleLysglulle 1008
QY 214 TAT-----ACGTATGTGAAGAAGCGTGATCAATCAACAAATGA--- 252
Db 1009 HisasnTyrGlnThrasnillethrTyrGlnThrLysasnlglnalavalThrProserCys 1028
QY 253 -----AAAACAGATATTATCATGATTCGACATATGATATATATGAT--- 297
Db 1029 CysTyrHisasnThrserHisilleleleuasnThrHISglu---AsnilleTyrGlnlgln 1047
QY 298 -----AATCATTTTACTTAAGACACATG-----CGTTAT 330
Db 1048 LysLysLysasnasnValleuLeuasnaspaspleuTyraspqlulleuGlnuarTyr 1067
QY 331 TTGATATATAT-----TATTTAAATCTGCAGAGAATAAT 342
Db 1068 MetasnLysIlelelaspnasleuphepheSerPheleleuLysValglyLysLys 1087
QY 343 ---ACACGCGAATAATATA-----TATTTAAATCTGCAGAGAATAAT 381
Db 1088 TyrThrHisTyrasnleuserSerSerleuileglTyrasnlylleuGlnlgln 1107
QY 382 GAACACACATTAATAAAGAGAGCGTACATTTTACGCCAAATAAATGATTTTAT 441
Db 1108 Lys-----LysaspLysThrIle-----AsnasnlglnasnaspIlelle 1120
QY 442 GAA-----AAACGTGTTAAATCTTCATCAATTTAGAT----- 474
Db 1121 LysIleaspnasnlnLysasnlglnLysnThrLysasnValaspnasnmetTyrThrserSer 1140
QY 475 ---TTAACAATGTTT-----TTAATGCTTTTAA----- 501
Db 1141 LysCysThrLysPheProPheasnilleasnLysaspLysLysTyrSerilleasnilleTyr 1160

```

QY 501 ----- 501
 Db 1161 PheLeuValTyrAspAsnIleLeuSerTyrAsnLysLysIleAsnLysGluGluIleGlu 1180
 QY 502 -----TTTAATATTTATGATTAACCTTANGAAACCAATACATTCATTCGACACA----- 549
 Db 1181 LysIleTyrAsnIleLeuAspAsnMetIleLysTyrLysGlnAsnValLeuThrGluAsp 1200
 QY 549 ----- 549
 Db 1201 AsnPhenTyrTyrIleIleSerAlaLeuLeuLysAlaGlnAsnPhenGluHisGluValTyr 1220
 QY 549 ----- 549
 Db 1221 LysMetTyrTyrGluTyrMetLysLysCysGlySerCysIleAsnIleLysTyrValPhe 1240
 QY 550 -----TTAGGTAGAAATTACT-----GATGGGTGTTATTACACAAATACACAACTAAACA 603
 Db 1241 PheIleMetLysArgIlePheGluAspThrProTyrIleThr-----TyrLysGln 1257
 QY 604 GATTTTAAATTATACGATTTTGTATTAAGATTAATGATGATGATAGT----- 651
 Db 1258 AspThrSerLeuAspIle---AspLysGluAsnIleLeuAsnAsnSerIleLysLysTyr 1276
 QY 652 -----GAACCTATGCTATGCTGTGAATGT-----TTTCAAAACTCACACCT 696
 Db 1277 AsnIleGlySerThrTyrTyrTyrAsnMetLysCysAspLysTyrGluLysCysAsnLys 1296
 QY 697 GAACAACTACATAC-----ATTCAATAAGACGTGATTAATATAGT----- 738
 Db 1297 TyrAsp---AsnTyrAspLysTyrAsnIleLeuAsnAspIleIleLysLeuSerGluGln 1315
 QY 739 -----ATGGCCATATTCATTTACTGATATATTTCCAAATTTTACTATACAAATTA 792
 Db 1316 IleIleLeuSerHisIleHisTyr-----IleLysAsnPhen 1327
 QY 793 ACATTTTCATGATATATATGATCTTACTGTAATATGAATGACACGTTTCAGTA 852
 Db 1328 ThrPheLeuLysGluValLeuHisThrTyrMetLys----- 1339
 QY 853 CTCACCAATATCATATTAATAATATCT-----TATACATATATCAT 897
 Db 1340 -----LysAspIleTyrIleLysCysTyrLeuPheTyrTyrProHisPheHis 1355
 QY 898 -----TTCCATGATGATGATTTTATGACAT----- 924
 Db 1356 AsnPhenValLeuThrTyrPheHisLysPheLeuThrHisAspGlnPheAsnLysAsnVal 1375
 QY 925 -----ATTAAATCATTTCTATCGTGTGTTTAATATGATATAC 963
 Db 1376 LeuValLeuLeuIleAsnAsnIleAlaSerPheTyr-----TyrThrLeuHisAsn 1392
 QY 964 ACCAATACATATAACAACATAATGATGACCTGTGTTTCTATGACATCATTCGAGT 1023
 Db 1393 AsnThrTyrThrSerSerTyrIleIleArg-----LysLysAspThrGlnArgGlu 1409
 QY 1024 TATCTTATGTCGATGATCATGATAAAATTCACACATGCTTATGACACACATAT 1083
 Db 1410 TyrGlnLysIleIleLysGluLysLysIle-----IleGluHisAsn 1423
 QY 1084 TCAGAACCAACG-----TTAATCCCTACTTTT-----TTAGATGAT 1119
 Db 1424 AsnGlnLysAsnLysGluLysLeuIleAsnHisTyrGluAspIleAsnIleLeuAspGlu 1443
 QY 1120 GACATATAT----- 1128
 Db 1444 GluAsnPhenLysGluAspHisLysAspIleLysValLeuLysLysTyrLysAsnGluTyr 1463
 QY 1129 -----TTTTCATTAATATAGATGATTAAGATGATTTTAACGATGATTTA 1173
 Db 1464 TyrTyrSerLysIlePheSerLeuTyrProLeuAspGlnIleHisLeuAsn----- 1480

QY 1174 TTAATTAATAATTAATACGCTGATATACGTCAATGATTTGTAATAATATATGAT 1233
 Db 1481 -----IleGluLeuLys-----GluGluMetValAlaLysAspLysThrAsnGln 1456
 QY 1234 AATGATATACGTTAATATCAATACAAATATACATTAAGATGATTCACAGCATTCGGGATTT 1293
 Db 1497 Gly-----AsnIleGlySerAsnLeuLeu-----LeuThrGlyAla 1508
 QY 1294 GATTCGATGCAATATACGCTGTTAATTCGTTGTTATATATGATGATGATTCATTCAGCA 1353
 Db 1509 -----Ser 1509
 QY 1354 CGTGATATTTATTTTCAAACTATTTATTAACACAGAGTAAGTAAAA-----AAC 1407
 Db 1510 LysAspIleThrSerTyrAsnTyrIleAspThrTyrIleLysMetGluLeuLys 1529
 QY 1408 AAATCATATATGACATACATCCTTACACATACATATCTATGATGATATCAACAGAACCA 1467
 Db 1530 LysLeuAsnIleLeuLeuProThrLeuTyr-----IleLysGluIleLysAsnLysSerPro 1548
 QY 1468 TACTCA----- 1473
 Db 1549 HisGluIleLysLeuSerSerMetAsnIleIleAspIlePheValSerLeuLysAsnVal 1568
 QY 1474 -----AATGAGAGCTATG-----TTATCTAAAGTCGTTTAAAGATTTATAT 1518
 Db 1569 LysIleArgAsnGluAspIleMetTyrLysLeuSerGlnLysTyrIleMetAspIlePhe 1588
 QY 1519 GGCATATCCGTCATATGCTTACATTTATTAATTCCTGTTATGATGATTAACAAAGACTA 1578
 Db 1589 -----PheHisAsnLysValLysLeuGluTyrGlnIleLysPhe 1602
 QY 1579 TACATATATCATTAACGTTTACAAAAAC-----ACTGACCTAATATATATTTCTACATTT 1635
 Db 1603 LeuAsnSerLeuThrPheLeuAspTyrIleLysGluAlaAspLeuPheLysThrPhe 1622
 QY 1636 GTCACATACGCTTCATG----- 1653
 Db 1623 PhePheLysLysAsnLysIleAsnLysIleGlnLysGluLysLysGlnAsnAsn 1642
 QY 1654 TATACCTATTTGGTTCCTTCATTAACGGAACGTAATGACAGCATTTATTT 1713
 Db 1643 TyrAsnLeuLeu-----TyrThrHisPheLeuLysIleProIle---HisAsnCysIle 1659
 QY 1714 TATTCGATACGATGATGTTGATATGAAATCCGTTGTAACCTTATGAAACCCAGT 1773
 Db 1660 TyrIleProAsnIleSerSerTyrIleLeuAsnHelle-----Ser 1673
 QY 1774 TTATTCGACCCGATAGCCCTTAGTAATGGATATTGAAAACGACAGATAGATAG 1833
 Db 1674 IleTyrAspTyrPhe-----GluLysLysAspGlnTyr 1684
 QY 1834 TTTTGATGATCATATTAAGAAATATGCAATGATGAATGGAAGATTAATAATTCCTCT 1893
 Db 1685 ValIle-----TyrLysLysLeuLeuTyr----- 1692
 QY 1894 GCTGGTATACGAAAAAGCCCTTTGATACAAAGCGTCGATTTTGAAACCTTTGACGTGA 1953
 Db 1692 ----- 1692
 QY 1954 CAATTCCTTTGACGGTCCATATTTGAAACATATAAAGTATCTATATGACGAAGTACA 2013
 Db 1693 ---PheLeuAspGluTyrLeuLysSerHisAsn-LysIleAsnSerMetAsnSerLeuAs 1711
 QY 2014 ATATCGA---TATATCCGTTAAACGAAATTTGATGCGTATGATATGATGATGAT 2070
 Db 1711 PlysArgAsnIleIleLeuIleIleIleLeuLeuTyrIleSerSerProLeuAsnIle 1731
 QY 2071 TTTACGATGACCTTAATATGAAAACGTAATTTATATTAAGAAGCTAGAGAAAATTC 2130
 Db 1731 eleuSerIleArgLeu----- 1736
 QY 2131 GACATATGTCATTTGATGATATCTTTATATGAAAGTGACATCGCTGCATTTT----- 2185


```

Db 1737 -----GlnThrLeuArgIlePheHyrTYrTYrIleIleGlnSerAsnTyrPheSerly 1754
QY 2186 -CACTTAAGACTTAT 2200
Db 1754 SHASnIleThrTyr 1759

RESULT 10
ID 096240 PRELIMINARY: PRT: 1182 AA.
AC 096240:
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE Predicted integral membrane protein.
GN PF0735C.
OS Plasmodium falciparum.
OC Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.
OX NCBI_TaxID=5833;

[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetelini H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perlea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AE001414; AAC71936.1;
SQ SEQUENCE 1182 AA; 144123 MW; 130BDB95717BFCA CRC64;

Alignment Scores:
Pred. No.: 3.56e-06 Length: 1182
Score: 229.50 Matches: 185
Percent Similarity: 35.96% Conservative: 135
Best Local Similarity: 20.79% Mismatches: 287
Query Match: 5.64% Indels: 283
DB: Gaps: 46

US-09-727-892a-2 (1-2286) x 096240 (1-1182)
QY 25 TATCATTAACATGACGTGCAATGATTTTATCTAGCGATATAGAACATTAAGCGTACAT 84
Db 141 PheAsnTyrSAsnSAsnArg-----IleTyrPheSAsnIleValLeuPheLysAsn 157
QY 85 AAAGTTAAGCGAGAAAAAACCACCAATATATAAAGCTTACTTCTGTAGCAATT 144
Db 158 AspLeuLeuGlnArgAsn-----IleAsnIleSerTyrGlnSerAsnIle 173
QY 145 GGTGGTTTAT-----GGTTATGAATGATGATGATTTTTCGAGATTCGATCT 198
Db 174 AspAsnMetSerArgGlnGlnGlnValHisHisLysArgSAspIleLeuIleAsnThrGlnCys 193
QY 199 TTTTATGAC-----GCATTTTATACGTATGTGAAGA--- 231
Db 194 LeuTyrAsnIleAsnAspLeuPheAlaLeuPheIlePheTyrValHisIleLysArgPhe 213
QY 232 -----CGTATACATCACAAATGAAAAACAAATATTAATCATGATGACCAT 279
Db 214 TyrPheAspPhePhePheThrIleLeuLysAsnIleAsnSAspMet-----Glu 229
QY 280 AACTGTATAATAATACGATTAATCATTTTCTACTTAAGACACCATCGCTTAATTTGATAAT 339
Db 230 SerThrAsnSAspTyrLysAsn-----ValCysTyrMetLysAsn 242
QY 340 ATTACAGCGCAAAATATATATTAATAATCTGCACAGAAAT----- 381
Db 243 IleHisLysGlnHisIleTyrHisIlePheProHisLysSAsnTyrTyrAsnIleGlnAsn 262
QY 382 -----GAACACACATTAATAAATGAAGAGCTACTATATTTAGCCAAAAAT----- 426

```

```

Db 263 MetAsnSerGluTyrCysLeuLysPheLeuLysAlaCysIleGlnLeuLysAsnIleIle 282
QY 427 CAATAATGTAAT-----TTAGAAAAACGTTTAAATCTTCATCAAT 468
Db 283 SerAsnIleValAsnIleAsnLysLysLysLysGlnLysAsnValThrAsnHisGlnAsn 302
QY 469 TTAGATTTAACAAATGTTTAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 528
Db 303 AsnIleArgThrCysArgIleAsnThrPheValPhe-----IleLysAsnAlaIlePhe 320
QY 529 ACCAATACATCAATGCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 588
Db 321 LysLysCysLysIle-----IleLysLysLys----- 329
QY 589 TCACAACCTTAATAACAGATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 648
Db 330 -----GluLysLysLysLysLysLysAsnAsp 337
QY 649 AGTGAAGCCTATGACATAGCTGTAAGATGTTTGCAAAACCTGACACCTGAACTTACA 708
Db 338 GluGlnIleTyrIleLysAla----- 344
QY 709 TACATTCATTAATGACGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 768
Db 345 TyrIleHisAsnSerVal-----TyrThrAsnIlePhe 355
QY 769 CCA-----AATTTGACTTATACAAATTAACATTTTCAATGAATTAATG 813
Db 356 LysAspMetLeuLeuHisAsnIleLysIleGluArgLysLysLysLysLysLysLysAsn 375
QY 814 GAATCTTACTGTAATTAATAAATGACACGCTTTGACGTAAGTCAACCAATATCAAGAT--- 870
Db 376 AsnLysIleIleAsnAsnLysIleIleAsnLysAsnIleIleGluLeuPheAsnSAsn 395
QY 871 ---ATTAATAATCTTATACATTAATCAATTC-----CATATATG 909
Db 396 IleIleArgLysLysLysLysIleHisPhePheLeuLysLysGlnLysTyrLysSAsn 415
QY 910 AATTATTATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 969
Db 416 ThrTyrHisLysPheLysLysArgLysAspMetAsnThrLeuIleMetCysAsp---Lys 434
QY 970 TACATTAACAACTAATTAATGAGCCTGT---TTTCTATTGACATCAATTCGCTAT 1026
Db 435 TyrIleAsnLysSerIle-----CysLeuPheLeuAsnAsnPheGlnAspSerSer 451
QY 1027 CCTTATGATGATGATGATGAAAAATTCACACATGTTTACTTTTACGACACATATCA 1086
Db 452 IlePheIleLysTyrMetLysIleIle----- 460
QY 1087 GAACCAACGTTAATCCCTACTTTTATGATGATGACAAATATTT-----TCATTAAT 1140
Db 461 LysIleAlaAsnIleIleAsnTyrLeuTyrAspAspHisValPheIleLysSerLeuMet 480
QY 1141 AAGATGATTAAGAT-----GATATTAACGATGATTAATTAATTAATTAATTA 1188
Db 481 LysCysValLysLysAsnCysAlaTyrPheThrGlnAspLeuLeuIleHyrLys 500
QY 1189 TCACGTGTA-----TTACGTCAATGATTTTAATTAATTAATTAAT 1227
Db 501 TrpLysThrHisMetAsnAsnLeuAspAsnIleAsnGlnHisAsnSAsnLysTyrLysAsn 520
QY 1228 AATGATTAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1269
Db 521 LysHisAsnSAsnMetTyrIleLysThrAspLysValLysAspAsnSAsnValLeuPhe 540
QY 1270 -----ATGATTCACGATTAACGCGATTAATGATTCGATG---CATATACGCTTAAT 1317
Db 541 ProPheSerLeuIleLysAsp-----AspIlePheArgHisIleGlu----- 554
QY 1318 TCGTTGTTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
Db 555 -----AspTyrHisPheHisHisIleLysAspIleIleTyrIleCysTyr 569

```

```

QY 1378 TTTATTTAAACACAGGTAAGTAAAAAATAATGACATTCACCTTACGACTAT 1437
Db 570 -----LysAsnLysLeu-----TyrGluTyr 576
QY 1438 CACATTCATGATATACAGACAGACCACTACTCAATGAGGAGGTATGTTACTATA 1497
Db 577 LysLeuPheHisLysIleIleAsnHisLeuIleAsnAsnIleAsnLysIleCysSerLys 596
QY 1497 ----- 1497
Db 597 TyrLeuValThrIleIleLeuLeuTyrAsnLysLeuAsnLysThrGlnLeuLys 616
QY 1498 -----GTGGTTTAAATGATATATAGCATACCTGACCTTACCTGACACT--- 1542
Db 617 GluLeuLeuPheIleLeuLeuAsnLysTyrArg---ProSerLeuLysGlnArgAsn 634
QY 1543 -----TTTAACTTATTCGGTTTATGATGATACATGACACTATAC--- 1581
Db 635 LysArgAsnAsnIleSerIleAsnAsnIleTyrLeuLysAsnIleAsnLysLysTyrIle 654
QY 1581 ----- 1581
Db 655 LysLysLysLysLysLysLysLysTyrIleTyrIleTyrThrIleCysLysLysAsn 674
QY 1582 AATATCATTAACGCTTACAAA-----AACACTGGAAGCTAATATATTA 1623
Db 675 AsnValGlyAsnIleHisLysHisAsnValMetMetThrSerAsnHisAsnHisLeu 694
QY 1624 TTCTCTACATTTGTCACATCAGCTCATGTAATCATATTGTTCCCTTCCAACTCTTA 1683
Db 695 PheArgSerPheGluTyrValLys---ValHisLysLeuLeuPheIleAsnIleLeu 713
QY 1684 ACGAAGATGAAATGACACATTTTATTTGCGATACGATGTTTGTATGAAA 1743
Db 714 IleLysSerAsnIleTyrIleAsnTyrGluTyr-----SerLeuTyrPheLeu 729
QY 1744 TCCGTTGTTAAA-----CCCTTATGAACCCCACTTATTCGACCCGATACCTTA 1794
Db 730 SerLeuIleLysGlnLysHisAlaPheIleLysLysGlyPheTyrIleLeuLysTyr 749
QY 1795 GGTAAATGGGATATGAAACGACAGATA-----GATAAGATGTTTGTATA 1839
Db 750 IleLeuPheHisIleGlnAsnHisIleIleTyrLysSerTyrGlnHisIlePheAsn 769
QY 1840 CTGAATCATAGAAATATGCAATATGAGTGAAGATTAATTA----- 1884
Db 770 ProTyrAsnLysTyrAsnIleTyrAsnIleTyrAsnIleLysCysThrLeuProGln 789
QY 1885 -----ATTGCTTCTGCTGGTATACCGAAA 1908
Db 790 IleLeuGlyThrSerAsnIleTyrSerLeuIleTyrValAlaPheLeuTyrSerThrAsn 809
QY 1909 AACGCGTTTGTACAGAGCGTATTTGAAACCTTTGACGGAACATTTCTTGACGCT 1968
Db 810 AsnThrIleAsnPheIleLysIlePhePheThrIleIle---GlnLysPheTyrAspSer 828
QY 1966 GCCATTT-----ATTGAAACAAATAAAGATATCTAATATGAGCAAGTACATATCG 2019
Db 839 SerMetIleLysGlnIleGlnAsnAspLysAsnLysTyrGlnHisIleSerCysHisAsn 848
QY 2020 ATATTCGCGTCTAAACGAAATGCTATGCTATGATATGATGAAATATTTACGAT 2079
Db 849 TyrSerProLysLysAspAsn-----SerGluTyrTyrIlePro 861
QY 2080 GAACCTAATATGAAACGTAATATATTAAGAACGCTAGAGAAATTTGACGACTAGT 2139
Db 862 AspAspHisAsnLysLeuLeuTyr-----AsnTyrSerTyrAsn 874
QY 2140 CAATTGATGATATTTCTTATATTTGAAGTACATCGGTTTCATTTTCACTTAAAGACTTA 2199
Db 875 GlnLeuTyrGluLysAsnHisPheAsnAspAsp-----AsnIlePheIleHisAspLeu 892

```

```

QY 2200 TTTCGAGTTGACGCTTACGATACATACAAA 2229
Db 893 LysIleTyrGluArgAsnIleAsnHisLys 902

```

RESULT 11

```

ID Q9XM75 PRELIMINARY; PRT; 1035 AA.
AC Q9XM75;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE DNA polymerase.
OS Neurospora crassa.
OC Mitochondrion. Plasmid Harbin-3.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARBIN-3993;
RA Griffiths A.J.F., Xu Y., Turitsa I.;
RT "Divergence of a linear and a circular plasmid in disjunct natural
RL isolates of the fungus Neurospora."
RL Plasmid 0:0-0(1999).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
+ {DNA}(n).
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL: AF133505; AD31446.1; -.
DR InterPro: IPR002064; DNA_Pol_B.
DR InterPro: IPR004868; DNA_Pol_B_2.
DR Pfam: PF03175; DNA_Pol_B_2; 1.
DR PRINTS: PR00106; DNAPOLE.
DR SMART: SM00486; POLBc; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase; plasmid.
SQ SEQUENCE 1035 AA; 120868 MW; C8565D2B61BDAF3D CnC64;

```

Alignment Scores:

```

Pred. No.: 6.67e-06 Length: 1035
Score: 225.00 Matches: 167
Percent Similarity: 34.57% Conservative: 132
Best Local Similarity: 19.31% Mismatches: 288
Query Match: 5.53% Indels: 278
DB: 8 Gaps: 40

```

US-09-727-892a-2 (1-2286) x Q9XM75 (1-1035)

```

QY 1 ATGGCATTTACTAGATGCATGCAATATATCAATGAACGCGAATGATTTTACTCG 60
Db 298 ValGlyAsnIleGluProThrLysArgAspLysGlnAspLysLysIleLeuAlaPhe 317
QY 61 GATATGAAACATTTAGCGTACATTAAGTAAACGACGAGAAAAACCAACCAATATMAA 120
Db 318 AspIleGluThrPheGlnValProThrGlyAsnGlyAspSerThrMetIleAlaTyr--- 336
QY 121 AACGTTACTATTCTGAGCAATGCTGGTTGTTAATGTTATGATATGATGTTGAAGTA 180
Db 337 -----AlaCysGlyPheTyrAspLysAsnLysSerLeuThrTyr 350
QY 181 TTTCGAGTTTCGAAATCTTTTATGACGCAATTTTATACGTATGTAAGAAAGAGCTGATACA 240
Db 351 IleSerAspPheIleSerGlnArgGluMetLeuAlaCysIleLys-----AspMet 368
QY 241 ATACAAAAATCAAAACAGATATATCATGATGTCACATTAAGCTATTAATACGATTAAT 300
Db 369 LeuLysTyrAspLysHisThrVal-----TyrCysHisAsnPheSerLysPheAspIle 386
QY 301 CATTTTTCATTAAGACACACATGCGTTATTTT-----GATAATTTTACACGCGAAAT 354
Db 387 AsnPheIleIleLysIleLeuValGlnGluPheValGlnLysIleIleSerLysAsp 406

```

QY	355	ATATATTTA-----	AAATGCGCAAA	375
		::: :::		
Db	407	LeuaspilleuSerlleYlsleSerTYrLysPheGluProLysLysLyselyLys		426
QY	376	GAAAATGACACACATTAAAAATGAAGAGCGC-----	ACT	411
		:::		
Db	427	AlaGluArgIshThrIleThrIleLaaspSerCyArgLeuProGlySerLeuasp		446
QY	412	ATTTTAGCCAAAATGCAAAATGTAAATTTTGAAGAAACGGTTAAATCTTCATCATTTTA		471
		:::		
Db	447	LysLeuAlaLysaspHisasnIleIleThrLysLys-----		458
QY	472	GATTTAACAATGTTTTTAAATGGTTTTTAAATTTAATTTATTTATGTAATCTTATGAAAAAC		531
Db	458	-----		458
QY	532	AATACATCAATTCGACACATAGGTAG-----	AAATTCCTGAGTGGTGTATTTA	582
Db	459	-----	-----	471
QY	583	ACAGAAATCACAACTTAAACAGATTTTAAATATATACATTTTGTATTAAGATAATGATATG		642
		:::		
Db	472	GluIyrValGlyLeuIleProAspTYrGluTYrIleaspProLysLysGlyGluMet		491
QY	643	AATATATGTGAAGCTATGACTATGCTGCGTAATGTTTTGCAAAACATCA-----	CCT	696
Db	492	-----IleThrIleuPheGluITrPAlaAlaMetTYrThrAsnLysTrpSerThrArgLys		509
QY	697	GAACAACTTCATCATTCATTCATTAATGACGTG-----	ATTAATATAGGTATG	741
		:::	:::	
Db	510	GluThrIleIleTYrLeuGluLysAspIleLysAlaLeuTYrGluLeuMetGluMet		529
QY	742	TGCCATATTCATTATGATGATATATTTCCAAATTTTGACATATACAAATTAACATTGTTCA		801
		:::	:::	
Db	530	SerAsnSerThrTYrSerThrPheArgIleAsnIleThrArgValLysThrAlaSerAla		549
QY	802	TTG-----	AAATATATGGAATCTTACTTGAAAT	828
			:::	
Db	550	LeuAlaPheLeuValTYrArgThrLeuPheLeuProAsnGluValGluGluGluGluGlu		569
QY	829	AATGAATGACACGT-----	TTTCAGTTACTCACCAATATACAGATATTTAA	876
			:::	
Db	570	AsnGluITrThrSerProAsnAsnIleLeuSerLeuPheAspLysLysGluGluLys		589
QY	877	ATATCTTATACAACTTATCATTTCCATGATATG-----	AATTTTATGACTATTTATAA	930
			:::	
Db	590	LeuITr-----ProLysTYrPheLeuProLysLeuLysGlyArgLeuGluArgAlaValArg		608
QY	931	TCATCTATCTCGGGGTGTTTAAATATGATATATACACCAAAATACATAAACAACAACTTATGAT		990
		:::		
Db	609	AlaAlaTYrPheGlyArgAsnGluIlePheIleProIleIleAsnAla-----		626
QY	991	GAGCCTTTTTCATTATGACATCAATTCAGATATCTCTTATGTGATGTATCATGAAAA		105
Db	627	-----PheSerPheAspPheAsnSerLeuTYrProThrAlaMet-----	Met	640
QY	1051	ATTCACAACTGATGTATCTTATAGCAACACTATTCAGAACCAAGCAAGCTTAAATCCCTACTTTT		111
		:::		
Db	641	MetProMet-----	ProValGlyIleProValHis	650
QY	1111	TTAGTATGATGACAAATTAATTTTCATTAATATAGATTGATAAGATGATTTTAACAGATGAT		117
		:::		
Db	651	ThrPheSerLysAsnLeuAsnGluIlePheGlyPhe-----		662
QY	1171	TTATTATTAATTAATTAATACAGGTATATAGTCAAAATGATTGTAAATACATAATATAT		1231
		::: :::		
Db	663	-----ValArgAlaLysIleIleThrProAlaIle-----		6721
QY	1231	GATTAATGATTACGTTAAATATCAATCAAAATACATTAGAATGATTCAGACATTAACGGGT		129
Db	673	-----	AsnIleProVal	676

QY	1291	ATGATTCGCATCCATATACGGTTTAACTCGTTGTATTATATAGCATGCAATCCTTCAT	1350
Db	677	LeupProcs--ArgVallyValasnIy-----	685
QY	1351	GCACGGATATTATTTTTCANAAC-----TATTTT-----ATT	1383
Db	686	ValGIInLyLeuIIeHerProIleGIyGIyPrIInGIyTrPTySerergInGIyLeu	705
QY	1384	AAAACAAGATGATTAATAAAAAACAATATCATATGACATGACACTTACGACATATGCATT	1443
Db	706	LysLeuAlaIValGIyGIyGIyTrPylLysIleGIyValLeuGIySerTyValPheGIyLys	725
QY	1444	ACTGATGAT-----ATCACACACACCCTACG	1470
Db	726	ArgAspAspProPheLysGLuTyrlIleGIyInHisPheIleSerIleLysaspSnITrLys	745
QY	1471	TCAAAATGACAGAGGTATGTATTCTAAGAAGTCGTTTTAAATGATATATAGCATACCTGCA	1530
Db	746	GlySerTyLysGIySmet---AlatLysLeuLeuasnIThrLeuTyrlGIyGIyTrnGIy	764
QY	1531	TTAGCTGCATTTTAACTTATTCGCTTAGATGATGAATACAAATGACATATACATTCATT	1590
Db	765	MetAsnAspSerAlaIaIGluIleLysMeLeuThIThrAsmLIuLeuaspSnIle----	783
QY	1591	AACGGTTACAAAACACATGACAGCTAATATATATTTCTCATATTGTC-----	1638
Db	784	-----GlnLeuThrAsnValIleHisGIyPheGIyValAspAspAspLys	799
QY	1639	-----ACATCACCTTCATTG	1653
Db	800	HISTyValArGIyAspLysLysProCysProValLeuCysAlaGlnSerGIyLysAsn	819
QY	1654	TATACCTATTGGTGCTCTTCATCTACCTATACGGAAGTGAATATACGACAATTTATT	1713
Db	820	TyrGIyLeuLeu-----SerTyrlLeu---AspGIyGIyLysaspSpGIyPheIle	835
QY	1713	-----	1713
Db	836	IleasnSerThrSerIleAlaIaIaIaIaIaSerITripSerArGIyLeuMetTyrlLys	855
QY	1714	-----TATTCGATGATGATAGTTTGATATGAATCCGTTGTATAA	1755
Db	856	HISlleIleasnSerAlaIaTyrlAspThrAspSerIlePhe-----ValGIyLys	872
QY	1756	CCCTTATTGAACCCAGCTTATTTCGACCCGATAGCCCTTACGTAATAGGATATTGAA---	1812
Db	873	ProLeu-----AspSerAlaPheIleGIyGIyGIyCysGIyLysPheLysalaGIyTyrl	890
QY	1813	---AACGACACAGATAGATAGATGATTGTCTACTGATCATAGAAATPGCATATGAGAGT	1869
Db	891	AsnGIyLynLeuIleLysArGIaIaIePheIleSerGIyLysLeuTyrlLeuLeuAspHe	910
QY	1870	AATGGAAGATTAATAATGTCTTCTGCTGATATACCAAAAACGCCCTTGATACAGCGTC	1929
Db	911	GIyGIyLysLeuGIyIleLysCysLysGIyIleThryLysasnLysaspasnIThrHis	930
QY	1930	GATTTTGAACCTTTGATCGACGACCAATTCCTTGACGGTGCATATTGAAAACATAAA	1989
Db	931	AsnLeuaspIleasnaspPheGIaIaIaLeuTyrlAsnGIy-----GlnSerArg	946
QY	1990	AGTATCTATATATAGCAA-----GCTACATATACATATATCCG	2028
Db	947	ValLeuPheGlnGIyArGIyPrgIyArgSerLeuGIyLeuGIyTrnValIThrValIySTyr	966
QY	2029	TCTTAAACTGCAATGTATATGCGTAAAGTATATATGATCAA-----	2067
Db	967	GlnLysTyrlAsnLeuIleSerGIy-----TyraspLysArGIyLysLeuTyrlSerLeu	984
QY	2068	-----TATTTACTGATGACTTAATATGAACGTGATTT-----	2103
Db	985	GIyLysITrpValasnIThrSerProLeuCysIleasnGIyasnPheGIyValIleSerLys	1004
QY	2104	---ATATTAAAGAAGCT-----AGACAAAATTCGACCATATGCAATTT	2145

```

Db 1005 AlaLeuValSerAspValGlyGluThrTrpTyrArgLysAsnGlyLeuHisTyrAsnLysTyr 1024
QY 2146 GATGATATCTCTTAT 2160
Db 1025 AsnHisIleIleTyr 1029

RESULT 12
ID 078938 PRELIMINARY; PRT; 571 AA.
AC 078938:
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE B type DNA polymerase.
GN kPOL.
OS Agrocyste aegerita.
OC Mitochondrion.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Boletiales; Agrocyste.
OX NCBI_TaxID=5400;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SM 47;
RX MEDLINE=99254810; PubMed=1032331;
RA Bois F., Barroso G., Gonzalez P., Labarere J.,
RT "Molecular cloning, sequence and expression of Aa-polB, a
RT mitochondrial gene encoding a family B DNA polymerase from the edible
RT basidiomycete Agrocyste aegerita."
RL Mol. Gen. Genet. 261:508-513(1999).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -> N DIPHOSPHATE
CC + (DNA)(N).
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL: AF061244; AAC3727.1;
DR InterPro: IPR002084; DNA_pol_B.
DR InterPro: IPR004688; DNA_pol_B.2.
DR Pfam: PF03175; DNA_pol_B_2; 1.
DR PRINTS: PR00106; DNAPOB.
DR SMART: SM00486; POLBc; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN_1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Mitochondrion.
SQ SEQUENCE 571 AA; 67014 MW; E647C22165F3911B CRC64;

Alignment Scores:
Pred. No.: 1.29e-05 Length: 571
Score: 220.50 Matches: 142
Percent Similarity: 36.63% Conservative: 110
Best Local Similarity: 20.64% Mismatches: 225
Query Match: 5.42% Indels: 211
Db: 8 Gaps: 37

US-09-727-892a-2 (1-2286) x 078938 (1-571)
QY 49 ATTTATACGGATAGCATTAAGCATTAAGTAAAGTAAACGACGAAAAAACCA 108
Db 14 lleIleSerMetAspLeuGluThrIleLeuIleAsn----- 26
QY 109 ACCAATATAAAGCTACTTATCTGTAGCAATGTTGGTTAATGTTATGAAT 168
Db 27 -----LysHisIleProTyr-----LeuLeuSerTrpTyrAspGly----- 38
QY 169 GATGTGAAGTATTCGAGTTTGAATCTTTATGCGCATTTTATACGTATGTCAA 228
Db 39 -----AsnIleThrLysSerTyr-----PheIleAspSerIleGluAsnIleGlu 54
QY 229 AGA-----CGTATACATCAACAATAAACAAGATATTATC----- 267
Db 55 AsnAsnIleGluAsnMetIleSerArgIleMetAsnAspIleCysIleArgLysTyrLys 74

```

```

QY 268 -----ATGATGTCACATACCTGTAATTAATACGATTAATCTTTTACTTAA 315
Db 75 AsnTyrLysValTyrIleuHisAsnPheAlaLysPheAspGlyTyrPheLeuValLysTyr 94
QY 316 ---GACACCATCGCTATTTGATATATTACACGGAGAAATATATTATAATCTGCA 372
Db 95 LeuSerLysLeuGlyPheIleAspAsnIle----- 104
QY 373 GAAGAAAATGAACACATTAATAAAGAGGCTACTATTTTAGCCAAAATCAAAAT 432
Db 104 ----- 104
QY 433 GTAATTTAGAAAAGCGTTTAATCTTCATCAATTTAGATTTACAAATGTTTAAAT 492
Db 105 lleIleAsnLysGlyArgIle-----IleThrLeuLysPheIleTyrAsn 119
QY 493 GGTATTAATTTAATATTATTTACTTATTAAGAAACCAATCAATTTGACAAACATA 552
Db 120 LysTyrSerIleThrPheLysAspSerTyrIleuLeuProSerLeuArgLysLeu 139
QY 553 GGTAGAATTTACTTGATGGTGGTTATTTAACAGATCAACACTTAATAACAGATTTAA 612
Db 140 CysLys-----SerPheAsnThrGlnThrGln-----LysAspIlePhePro 153
QY 613 TATACGATTTTGGATTAAGATATGATGAAT---GATAGTGAAGCCTATGACTATGCT 669
Db 154 TyrIleuLeu-----AspAspIleAsnTyrIleGlyGluValProAspTyr--- 168
QY 670 GTGAATGTTTTCGAAACCTACACCTGACACTTACATCACTTCAATATGACGATTT 729
Db 169 ---LysTyrPheCysAsnLeuGluMetGluIuTyrAsnAsnTyrLysSerAsnPheLys 187
QY 730 ATA-----TTAGCTATGTCGATTTATTCATTTATAGATGATATTT 768
Db 188 ValTrpAsnPheArgGluGluAlaIleLysTyrCysAsnLeuAspCysIleSerLeuTyr 207
QY 769 CCA---AATTTTGACTATACAAATTAACATTTTCATTTGAATATTATGAAATCTTACTG 825
Db 208 GluIleLeuTyrLysPheAsnThrLeuValPheAsn----- 219
QY 826 AATATGAAATGACACGCTTTCAGTTA---CTGACCAATATCAAGATATT--- 873
Db 220 -----LysPheGluLeuAsnIleAsnLysTyrProThrLeuProSerLeu 234
QY 874 -----AAAAATCTTAT-----ACACATTAATCATTCATGATATG 909
Db 235 SerPheAlaLeuPheLysThrLysTyrLeuLysGluAsnGluValHisMetLeuSerGly 254
QY 910 AATTTTATGACTATTAATATCATTTCTATCGTGCTGCT---TTAAATATGTTATACACC 966
Db 255 SerIleAlaThrAsnIleArgLysSerTyrThrGlyGlySerValAspMetTyr---Ile 273
QY 967 AATATCAATAAACAATAATGATGACCTGTTTTCCTATTTGACATCAATTCGAGTTAT 1026
Db 274 ProIleuIleLys-----AspSerLysIlePheIleTyrAspIleAsnSerLeuTyr 291
QY 1027 CCTATGTCATGTCATGAAAAAATTCACAATGTTATTAATCTTTTACGAACACTATCA 1086
Db 292 ProPheSerMetLysSerPheLysPheProIleGly----- 303
QY 1087 GAACCAACGTTAATCCCTACTTTTATAGATGACGACATATTTTCATATATTAAGATT 1146
Db 304 -----AsnProThrPhePheLysGlyAsp-----IleThrArgIle 315
QY 1147 GATTAAGATTAATTTAAGATGATTAATTAATTAATAATTAATCAAGCTATTAACGTCAA 1206
Db 316 AsnLysAspAlaPheGlyPhePheTyrCysLysIleIleThrProGluTyrLeuGluHis 335
QY 1207 ATGATGTGAATAATACATATATATGATTAATGATTAATGATTAATATCAATATACATA 1266
Db 336 ProIleuLeuIleThrHisLeuLysThrSerAspGlyIle-----ArgThrLeu 351
QY 1267 AGAATGATT-----CAAGACATTAAGGGATTTGATTTGACATATAGCTGTTAAT 1317

```


QY	874	AAATATCTC-----TATACACATTTAT---CATTTCCATGATATGAATTTT	913
Db	794	AGTtLeCysLeuAlaArgAlaLeuArgtLuhstLrlehsLwehstLysLeu---Cys	812
QY	916	TATGACTATATTAAATCATCTGATCTGATCGTGGTGTAAATATGTATACACCAATACATA	975
Db	813	ThAsPtyrGluLysLysLeuIleGlnPro-----	822
QY	976	AACAAACTAATATGAGACCGCTGTTTTCATGAGACATCAATCCGATTTAT-----	1026
Db	823	AsnGluIleLeuAspLysAspLeuIleAsnLysAsnIleSerSerLyrAsnAsnLys	842
QY	1027	-----CCTTATGATGATATCATGAAAATAATCCAAACATGATGTTATACCTTTACGAAAC	1080
Db	843	LysSerLysLeuValAsnTyrAsn-----IlePro-----PheAsnGluAsn	856
QY	1081	TATTCAGAACCAACGACTTAATCCCTACTTTTTCAGATGACACATATTATTCATTTAT	1140
Db	857	TyrLeuGlnLysCysLeuMet-----AspAspAsnAsnMetLyrLeuLyr	871
QY	1141	AAAGATTGAT-----AAAGATGATTATTAAC---	1164
Db	872	LeuLeuAspAspIlePheThrSerLeuAspProSerIleSerLysLysIlePheSerAsn	891
QY	1164	-----	1164
Db	892	LeuPheCysLysGluAspAsnIleSerPheLysAspAsnCysSerPheIleIleSerMet	911
QY	1165	-----GATGATTTATTAATAAATTAATACAGT	1194
Db	912	AsnLysSerThrLeuAspAsnPheLeuIleGlnAspIleLeuAspAsnValGlnTyrGlu	931
QY	1195	GTA-----TTCAGCAAAAGATGTTAAATCTATTAATATGATTAATGATATAC	1242
Db	932	ValAsnIlePheGluIleGlnAspLysThrLeuLysTyrArgGluAsnIleSerGluLyr	951
QY	1243	GTT-----AATATCAAT	1254
Db	952	MetGluLysAsnAsnLeuAsnIleThrLysGluSerHisTrpGluTyrSerAsnLeuAsn	971
QY	1255	ACAATATACATTAAGAATGATTCACAAGCATACGCGATGATTCGATCATATACGTTT	1314
Db	972	ThrIleAspLyrThrArgIleLysLeuPheAspGluValGluLeuAsnHisValLysHis	991
QY	1315	AATTCGTTTGTTATATGATGATGCAATCTT-----CATGCAAGTGATATTATT	1365
Db	992	SerAsnLysMetLleLyrLysGluAlaTyrPheValLysGluLysThrGluSerValSer	1011
QY	1366	TTTGAA-----AACAATTATTAATAACACAAAGATTAATAAACAAA	1410
Db	1012	PheGluIleAspSerIleAsnLysGluLyrIleLys-----LysMetLysLysLys	1028
QY	1411	ATCAATATGACATTCACCTTACGACATTCACATTTACTGATGATATCAACGAACCCATAC	1470
Db	1029	AsnTyrLysLysGluHisMetAsnLysAsnAsnLysAspAsnAsnAsnAsnAsn	1048
QY	1471	TCAATGAGGAGGTATATGTTATCTAAAGTCGTTTAAATGAGATTATAGCATACCTGCA	1530
Db	1049	SerAsnLysAsp-----	1052
QY	1531	TTAGCTCACATTTTAACATTATTCGCTTTAKATGTAATCAATGAACATATACATATTCAT	1590
Db	1053	-----AspHisIleAsnIle---AsnMetAsnAspAsnHisAspAsnLyrAsnAspIle	1069
QY	1591	AAC-----GGTACACAAAACAGCTAGCATAATATATATCTCTACATTTGTCACATCGT	1647
Db	1070	AsnLeuGluLysProAsnSerThrAspSer-----ProThrValSer	1083
QY	1648	TCATGTATATACTTATGTGGTTCCTTCCATATCTTAAACGAAGTGAAATTTAGACACAT	1707
Db	1084	SerLeuLysAsn-----GluTyrThrLeuAspThrLyrThrSerAsnAsn	1098

QY	1708	TTT-----	ATTATATGCATACCTGATAGTGTTCATATGAAA	174
Db	1099	SerAspLysGLuGLuIleValLysProLeuTyrLysAspThrHisGLuGLuPreAsnLys	1116	
QY	1744	TCCGCTGTTTAAACCCCTATTATGACCCAGTTTATTCACCCGATAGCCTTAGGTAAATGG	1803	
Db	1119	SerSerSerMetProPheValLysSerSer-----	1122	
QY	1804	GATATTGAAAACGACACAGATGATGAATGTTGTTCATGATCATATGAAGAAATATTCATAT	1866	
Db	1129	-----SerAsnMetIleAsnAsnProSerAsnPheLysTyr	1144	
QY	1864	GAAAGTGAATGGAAAGATTAAATATGCTTGTCTGTGTATACCGAAAACCCGCTTGTATACA	1923	
Db	1141	GLuAspAsn-----SerSerSerPheLysGly	1145	
QY	1924	AGCGTCGATTTTGAACACCTTTGATGCTGAACAATTCCTTATGC-----GGTCCCAATTAT	1977	
Db	1150	SerIleSerLeuGluThrTyrLeu-----TrrPhePheGlnGlnValGlyPheValLeu	1167	
QY	1978	GAATAACATTAAGTATCTCTATTAATGACCAAGTACATATTCGATATATCCGCTTAACT	2033	
Db	1168	LeuThrSerValIleIlePheMet-----LeuIleSerIlePheThrAsp-GLuI1	1184	
QY	2038	GAATTTGATGTGCTGAATGATATGATGATGAATTTTACTGATGATCACTTAATATGAACGT	2097	
Db	1184	elys---PheValPheLeuThrMetSerIleIleSerLysAsn-----	1198	
QY	2098	GAATTTATATTAAAGACGCTAGAGAAATTTGCACCATATGCAAT-----	2143	
Db	1199	-----AsnLysGLuHisSerAspThrIleLeuGlnLysGlnValArgTyr	1213	
QY	2144	-TTGATGATATTCTTTTATTATGAAGATGACGTGCTCAATTTCACTTAACACATTATTT	2202	
Db	1213	rLeuGLuTyrPheValIleLeuThrIleLeuThrIleSerLeu-----	1225	
QY	2203	CCAATTGACGCTTCAGTACATAACAAATCTGATTGTC-----ATATATTAAAA	2250	
Db	1226	-----ValThrSerGlyIleCysPheSerMetIleIleTyrGlyAs	1239	
QY	2251	CGTG 2254		
Db	1239	nile 1240		
RESULT 14				
ID	Q96206	PRELIMINARY;	PRT;	578 AA.
AC	Q96206;			
DT	01-FEB-1997 (TrEMBLrel. 02, Created)			
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	DNA polymerase.			
OS	Bacteriophage GA.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Leviviridae;			
CC	Levivirus.			
CC	NCBI_TaxID=12018;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Salas M.;			
RL	Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMATIC ACTIVITIES: DNA			
CC	SYNTHESIS (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT			
CC	DEGRADERS SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION (BY			
CC	SIMILARITY).			
CC	- 1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE			
CC	+ (DNA)(N).			
CC	- 2- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.			
DR	EMBL: X69687; CAA65712.1; -			
DR	InterPro: IPR002064; DNA_pol_B.			
DR	InterPro: IPR004868; DNA_pol_B.2.			
DR	Pfam: PF03175; DNA_pol_B_2; 1.			
DR	PRINTS: PR00106; DNAPOLB.			
DR	SMART: SM00486; POLBc; 1.			
R	SMART: SM00486; POLBc; 1.			

DR PROSITE: PS00116; DNA_POLYMERASE-B; UNKNOWN.1.
 KW DNA replication; DNA-binding; DNA-directed DNA polymerase.
 SQ SEQUENCE 578 AA; 67141 MW; 91465BAC07EE3A CRC64;

Alignment Scores:

Pred. No.:	1.59e-05	Length:	578
Score:	219.00	Matches:	138
Percent Similarity:	35.35%	Conservative:	107
Best Local Similarity:	19.91%	Mismatches:	248
Query Match:	5.38%	Indels:	200
DB:	9	Gaps:	34

US-09-727-892a-2 (1-2286) x Q96206 (1-578)

```

QY 139 GCAATGCTGGTTAATGATGTAATGATGTTGAGTATTCGAGTTTGGAACT 198
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 AATpGlyTrpMetAspIleTyrAsnThrAspLysTrp-----SerTyrGly 39

QY 199 TTTTATGACGATTTTATACGATGTGAAAGACGTGTACATACACAAATCAAAACA 258
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40 AspIleAspSerPheMetGluTrpAla-----LeuAsnSerAspSer 53

QY 259 GATATATCATGATGACGATACGATGATATTAATACGATATCATTTTCTTAAAGAC 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54 AspIleTyrPhe-----HisAsnLeu---LysPheAspGlySerPheIleLeuProTrp 70

QY 319 ACCATGCGTATTTGATATATTTACACGCAAAATATATTTTAAATCTGCAGAAAGA 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71 TrpLeuAlaGy-----AsnGlyTyrValHisThrGluGluAsp 82

QY 379 AATGACACACATTAATAATGAAAGCGCTACTATTATTCGCAAAATACAA----- 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 ArgThrAsnThrProLysGluPheThrThrIleSerGlyMetGlyLysTrpTyrAla 102

QY 430 ---AATGTAATTTAGAAAAAGCGTGTAAATCTTCAATCACTTTAGATTACAAATGTT 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 ValAspValCysIleAsnThrArgGlyLysAsnHisValAlaPheTyrAspSer 122

QY 487 TTAATGCTTTTAAATTTAATTTATTTGATTAACCTTTATGAACCAATCATCATTTGA 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 LeuLysLysLeuProPheLysVal----- 130

QY 547 ACATTAAGTAAGAAATTTACTTGTATGCTGTTATTTAACAGATCACAACCTTAAACAGAT 606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 -----GluHisIleAlaLysGly 136

QY 607 TTTAATATACGATTTTGTATAGATATATGATATGATGATGAGCCTATGACTAT 666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137 PheGlyLeuProValLeuLysGlyAspIleAspTyrLysLysTyrArgProValGlyTyr 156

QY 667 GCTGTAATGCTTTTGCAAAACTCACACCTGACCACTTACATCATTCATCAATGAGCTG 726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 157 Val-----MetAspAspAsnGluIleGluTyrLeuLysHisAspLeu 170

QY 727 ATTATATAGTATGTCGATATTCATATCATATATAGTATATTTTCCAAATTTAGCTATAAC 786
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 171 LeuIleValAlaLeuAla-----LeuArgSerMetPheAspAsn---AspPheThr 186

QY 787 AATTAACATTTTCATGATATTTATGTAACCTTACTTGATTAATGAATATACAGCTTT 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 SerMetThrValGlySerSerPalaLeuAsnThrTyr-----LysGluMetLeuGlyVal 204

QY 847 CAGTACTACCAACATATCAAGATATTAATAATATCTTATACATATATCATTTCCAGAT 906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 LysGluTrpGluLysTyr-----PheProValLeuSer 215

QY 907 ATGAAATTTTATGACTATTTAATCATCTTATCTGCTGTTTAAATATGATATAACACC 966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 LeuLysValAsnSerGluIleArgLysAlaTyrLysGlyGlyPheThrTyrValAsnPro 235

QY 967 AATATCATTAACAACTAATATGATGAGCCTTGTCTTCTATGATCATCAATTCAGATAT 1026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 LysTyrGluGlyGluThrVal---TyrGlyGlyMetValPheAspValAsnSerMetTyr 254
  
```

```

QY 1027 CCTATGATGATGATCAATGAAAAAATTCACATGCTTATCTTTACGACACGATTTCA 1086
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 ProIleMetCysIleLysAsnLysLeuLeuPro-----TyrGly 266

QY 1087 GAACCAACGTTATATC-----CTTACTTTTTCAGT 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 267 GluProValMetPheLysGlyGluTyrLysLysAsnValGluTyrProLeuTyrIleGln 286

QY 1117 GATGACAATTTATTTTCATATATAGATTTGATTAAGATGATGATTAACGATGATTATTA 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 287 GluValArgCysPhe-----PheGluLeuLysLysAspLysIlePro-----Cys 301

QY 1177 ATTAATTAATTAATCAAGCTGATATACGTAATGTAATAACTATATATATGAT----- 1233
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 IleGlnIleLysGlyAsn-----AlaArgPheGlyGluAsnGluLys 315

QY 1234 -----AATGATTAAGTTAATATATCAATACAAATACATTAAGATATGATCA 1278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 LeuSerThrSerGlyAspLysLysValAspLeuTyr----- 327

QY 1279 GACATTAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 ---ValThrAspValAspIleProLysLysLysLysLysLysLysLysLysLysLys 345

QY 1339 GATTAATTTATGACGATGATATTTATTTTCAAAATTTTATTAACAAAGGTTAG 1398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 346 GluPheIleGlyGly-----PheMetPheLysGlyPhe----- 356

QY 1399 TTAATAAACAATAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 -----IleGlyPhePheAspGluTyrIleAspArgPheMetGluIleLysAsn 372

QY 1459 GAACACCCATCTCAATATAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1518
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 373 SerProAspSerSerIleGluInsSerLeuGlnAlaLysLeuMetLeuAsnSerLeuTyr 392

QY 1519 GGC-----ATACCTGATTAACGTTACATTTT 1545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 GlyLysPheAlaThrAsnProAspIleThrGlyLysValProTyrLeu----- 408

QY 1546 AACTTATTCGCTTATGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1605
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 -----AspGluAsnGlyValLeuLysPheArgLysGly---GluLeu 421

QY 1606 ACTGACGTAATATATATATCTCT-----ACATTTGTCATCAAGCTTCAATGAT 1656
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 LysGluArgAspProValTyrThrProMetGlyCysPheIleThrAlaTyrAlaArgGlu 441

QY 1657 AACTTATTCGCTTATGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1716
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 442 AsnIleLeuSerAsnAlaIn-----LysLeuTyrProArgPheIleTyr 456

QY 1717 TCGATATGATATGATTTGATATGAAATCCGTTTAAACCCCTTATTAACCCAGTTTA 1776
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 457 AlaAspThrAspSerIleHisValGluGlyLeuGlyGluValAspAlaIleLysAspVal 476

QY 1777 TTGACCCCATAGCCCTAGTAAATGGAATATGGAAGCAACATATGATATGATGAT 1836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 477 IleAspProLysLysLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 1856

QY 1837 GTAACGTAATCATAGAAATATGATATGAAAGT-----AATGGAAG 1878
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 TyrValArgGlnLysThrTyrPheIleGluThrThrTyrPlysGluAsnAspLysGlyLys 515

QY 1879 ATT-----AAATGCTTTCGCTGCT 1899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 516 LeuValValCysGluProGlnAspAlaThrLysValLysProLysIleAlaCysLysIle 535

QY 1900 ATACGCAAAAGCCCTTGTATCAAGCGTGAATTTGAACCTTGTACGGAACAATTC 1959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 536 Met---SerAspAlaIleLysGluThrGlyLeuArgPheAsnGluPhe----- 549
  
```

QY 1960 TTGACGGTCCATTATTTGAAACAAATAGTCTATATGACGACGATACATATCG 2019
 Db 550 -----LyslelelytyrserThrsiglyserleu----- 559
 QY 2020 ATATATCCGTCTAAAGTAAATGTATGCTGATATGTA 2058
 Db 560 -----LysProlysaasnvalleuclglylval 568
 RESULT 15
 Q9EMR4 PRELIMINARY; PRT; 872 AA.
 AC Q9EMR4: 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE AMY135.
 GN AMY135.
 OS Amsacta moorei entomopoxvirus (AMEPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
 OC Entomopoxvirus B.
 OX NCBI_TaxID=28321;
 RN [1]
 RP MEDLINE=20396580; Pubmed=10936094;
 RX Bauman A.L., Glassberg K.J., Digians J., Shaw R., Farmerie W.,
 RA "Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:
 RT Analysis and Comparison with Other Poxviruses."
 RL Virology 274:120-139(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bauman A.L., Glassberg K.J., Digians J., Shaw R., Farmerie W.,
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF250284; AAG02841.1;
 DR InterPro: IPR004971; Pox_MCEL.
 DR Pfam: PF03291; Pox_MCEL; 1.
 SQ SEQUENCE 872 AA; 103047 MW; 20DB06798F9BBC44 CRC64;
 Alignment Scores:
 Pred. No.: 1.65e-05 Length: 872
 Score: 218.50 Matches: 18
 Percent Similarity: 36.04% Conservative: 131
 Best Local Similarity: 21.69% Mismatches: 279
 Query Match: 5.37% Indels: 305
 DB: 12 Gaps: 54
 US-09-727-892a-2 (1-2286) x Q9EMR4 (1-872)
 QY 82 AATAAGTTAAGCAGCAAAAAAACAACAATAATATATAAAGCTTATCTGATGACA 141
 Db 67 Asnlyslsyrserlyslsleuarglnlystlyrlyspyrprothrphgelnleala 86
 QY 142 ATGGTGGTTTATGGTTATGAATGATGATGATGATGATGATGATGATGATGATGAT 201
 Db 87 -----Asnserlyr-----phe 90
 QY 202 TATGACGATTTTATACGTATGTAAGACGT-----GATACATTCACA 246
 Db 91 Leuaspysleuthrasnasntrpqluarglysthrilleleserqluspsylleasn 110
 QY 247 AAAACAAAGACATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 306
 Db 111 Ilesnlysaasnlyrilleleuarglnlysthrilleleserqluspsylleasn 130
 QY 307 -----TTACTTAAGACACCATGCGTTATTTGATATATTTACACGCAAAATATA 357
 Db 131 Gluleuproleuasnaspilleleuaspyslileasnvalleuhevalseerqlleu 150
 QY 358 TATTTAAATCTGCAGAGAAATGAACACATTAATAATGAAGAGCTACTATTTTA 417
 Db 151 Tyr-----lle 152

QY 418 GCCAAATATCAATGTAATTTTGAAAAACGGTTAATCTCAATCAAT---TTAGAT 474
 Db 153 Ilelleasnaspilleuileysvalgluphelyslleysersnleleuilegylproleuser 172
 QY 475 TTACAAATGTTTAAATGCTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 534
 Db 173 SerasnlyslsleuleuserThrsPheasnaspilleuileuThryrarglysaenlle 192
 QY 535 ACA-----TCATTTGCAACATTTAGTGAACAAATTTACTGATGGTATTTAACA 585
 Db 193 ThrlyrlyleuagluilleuvalleuserThrsleuasnaspasnvalleuThyr 212
 QY 586 GAATCACAACCTTAAACAGATTTTAATTAAT----- 615
 Db 213 Aspsnleuvallysser---Pheglutlyrillelyrlysserlysaenlle 231
 QY 616 -----ACGATTTTGTATTAAGAT-----AATGAT 639
 Db 232 SerleuvalThrilleysasnlyslsProlyslleysthrisnlelleleuThryrlysa 251
 QY 640 ATGAAT-----GATAGTGAAGCCTATGATGCTGTGAATGTTTGCACAAATCACA 693
 Db 252 LeuasnThrleaspysgluserlyrilleuvalleleuys----- 265
 QY 694 CCTGACACACTTACATCATTTCAATGACGCG-----ATTATATTAGGTATG 741
 Db 266 -----Lleaspglyaspvalvalglupheasnvalmetasnlylle 279
 QY 742 TGCCAT---ATTCAATTAATAGATATATTTCCAAATTTTGAC----- 780
 Db 280 CysasnlellelelyrlyspmetvallyrlysaenPheSerCysasnleaspysasn 299
 QY 780 ----- 780
 Db 300 IleleuThrilleleuilegylmetglyglutlyrilleysvalaspasnvallyslslyrPro 319
 QY 781 -----TATAACAATTAATCAATTTTCATTAATTAATTAATTAATTAATTAATTAAT 834
 Db 320 PheThryrPheSerlyslsleuserlyrlysaen-----Asnlyslslelleasnlylle 336
 QY 835 ATGACAGCTTTTTCAGTTACTGACCAACATTAACAAT----- 870
 Db 337 Leuaspyslyrlyslsleuarglnlystlyrlyspyrprothrphgelnleala 396
 QY 871 ATTAATAATCTTTPACACAT-----TATCATTCATGATGATGATGATGATGATGATGAT 921
 Db 357 MetglnleuasphegluasnlyslsleuThrleuThrleuThrleuThrleuThrleuThr 376
 QY 922 TATATTAATCATCTCTCGT-----GGTGGTTAATATGATTAACACC 966
 Db 377 AsnvalleuThryrlyslsleuileglnasnserSerPheylsasnlelelyrlyspgly 396
 QY 967 AAATACATTAACAACATTAATGATGACCTGTTTTCATGATGATGATGATGATGATGAT 1026
 Db 397 Ilevalleuaspillethrlnaspasn-----Asperlyslsaspylr 411
 QY 1027 CCTATGTGATGTATCATGA-----AAATTCACACATGCTTATACCTT 1071
 Db 412 LysPheylsleasnspasnThrvalaspvalleuLysleuasnspThrlyrarglythr 431
 QY 1072 TACCAACACTTATGACAAACAGCTTATCTACTATTTTATGATGATGATGATGATGAT 1131
 Db 432 TyrilleuThrasnaspasnlyslsleuThr-----Phe 444
 QY 1132 TCATTAATTAAGATGCTTAAGATGTTTAAACGATTTTAAATTAATTAATTAATTAAT 1191
 Db 445 ThrleuThryrlyslsleuaspasnlyslsPheThrleuThryrlyslsleuThryrly 464
 QY 1192 CGTGTATTAACCTCAATGATTTGAATAATATATTAATTAATGATGATGATGATGATGAT 1251
 Db 465 Glulelleleu-----Tyrasnasn-----Tyrvalasnleu 475
 QY 1252 -----AATCAAAATACATTAAGACATGATCAAGACATTCGCGGTATGATTCG 1299


```

Db 476 LeuIlePheAsnAsnAsnAsnLysPheGlyProLysLysMetLeuSerProIleThrCys 495
QY 1300 -----ATGCATATACGCTGT-----AAT 1317
Db 496 IleValGluTyrSerPheLeuGluSerLysIleIleGlyLeuArgIleAspLysThrAsn 515
QY 1318 TCGTT-----CTTATATATGAAATGAAATAC 1344
Db 516 AsnPheTyrArgGlnAsnTyrAsnGlyAsnAsnLeuAspValIleLeuThrSerLysHis 535
QY 1345 TTTCATGCACGTGATATATTTTCAAACTATTTTATTAACACACAGTAAGTTAAA 1404
Db 536 IleHis-----GluGluPheProSerAsnTyrAsnIleAspTyrLeuMetSerLeuAsn 553
QY 1405 AACAAATCAATATACATCATCCTTACGACATGCATTTACTGTATGATATCAACAGAACAC 1464
Db 554 GluThrIleAsnVal-----IleAspAsnAsn 562
QY 1465 CCATATCTCAATAGAGAGATTATGTATCTAAAGTCGTTTAAATGATATATATGCATA 1524
Db 563 ProHisArg-----SerLysLeuLeuAsn----- 571
QY 1525 CCGCATTTACGTTACATTTTAACTTATTCGTTAGATGATACATGAACTA----- 1578
Db 572 -----LysGluValAsnLysTyrPheMet-----AsnAsnThrIleArgThr 585
QY 1579 ---TACATATCATTAACGTTTACAAAACACAGTACAGT---AATATATTTCTCTACA 1632
Db 586 SerIleAsnIleLeuThrAsnTyrLeuLysThrAsnGlyIleSerMetAlaIleSerLys 605
QY 1633 TTGTGTACA-----TCACGTTCAATGTATTAACCTA----- 1662
Db 606 LeuValIleThrIleProAsnArgTyrValLeuSerIleAspIleGlyArgGlyAsp 625
QY 1663 TTGGTTCCTTCCATACCTTA-----ACGAAAGTGAAT--- 1698
Db 626 LeuThrLysTyrTyrTyrValGlyIleThrGlyMetLeuGlyThrAspProAspIlePhe 645
QY 1699 -----GACGACAAATTTTATTTGCGATCTGATAGTTTGTATATGAAA 1743
Db 646 AlaIleLysGluAlaArgAspArgTyrLysLysLeuGlnThrIleSerAsnAlaGluAla 665
QY 1744 TCCGTTGTATAA---CCCTTATGTGAACCCAGTTTATTCGACCCGATAGCCTTAGGTAAA 1800
Db 666 SerIleTyrLysPheAspSerLeuAsnMetSerIleLeuAsnAsp----- 680
QY 1801 TGGGATATTGAAAGCAAGACAGATAGATAGTGTGTACTGAATCATAG----- 1851
Db 681 ---AsnTyrGluAsnGlu---IleLysAsnLysPheMetThrHisIleLysIleGlnTyr 698
QY 1852 -----AAATATGCATATGAAGTGAAGTGAAGS--- 1878
Db 699 PheGlyValIleGluTyrGluLeuAlaIleHisTyrSerTyrAsnAsnAsnThrLysAsp 718
QY 1879 -----ATTAATTTGCTTCTGCTGTATACCGAAAAACGCTTGTATACAGACGCTC 1929
Db 719 MetIleLeuLeuLysLeu-----LysAsnLeuSerAsnAspLys--- 731
QY 1930 GATTTTGAACCTTGTACGTCGACACATTTCTTGACGCTGCCATTA--- 1977
Db 732 -----ThrLysValIleIleThrCysLeuAspGlyAspGluIleThrAsnArgLeu 748
QY 1978 ---GAACACATATAAAGTATCTATATAGCAAGTACATATGCATATATCCGCTTAAA 2034
Db 749 AsnGluAsnProAsnLeuIleTyrAsnIleGlnProGlyIleThrTyrLysIleSerLys 768
QY 2035 ---ACTGAATTTGTATGTGTAATATATATGATATATTTACTGATGAACTTAATATAG 2091
Db 769 IleSerAspAspLysIleSerValLeuTyrAsnAlaIleMetThrGluTyrLeu----- 786
QY 2092 AAACGGAATTTATATATAAGAC-----GCTAGAGAAATTTTC--- 2130

```

```

Db 787 ---GluGluTyrIleIleThrAspLysIleIleAspAspPheIleMetTyrAsnPheIle 805
QY 2131 -----GACCATAGTCATTTGATGATATTTTATATATGAAAGTGCATACGCTTCATTT 2184
Db 806 LeuSerAspValCysLysPheAspAspIlePheLysTyrAsnSerAspLysSerValGlu 825
QY 2185 TCACIT---AACGACTTATTTCCAGTTGAACGTTCCAGTACATAC-----AATCTGAT 2235
Db 826 ValLeuSerAsnPheLeuArgLysSerThrTyrLysPheTyrAsnAspIleLysAsnAsp 845
QY 2236 TTGCATATATATAAAGCTGAACATGATGAATAATRAAAA 2274
Db 846 LysAsnIleTyr-----AsnAsnAspAspIleLysLys 856

```

Search completed: January 8, 2003, 17:06:12
Job time : 115 secs

This Page Blank (uspto)